HIV Database Workshop

www.hiv.lanl.gov
seq-info@lanl.gov

Presenters: Brian Foley & Karina Yusim

Database PIs: Bette Korber, Thomas Leitner, Karina Yusim

Additional database staff: Werner Abfalterer, Will Fischer, Peter Hraber, Elisabeth Sharon Fung, Robert Funkhouser, Kumkum Ganguly, Jenni Macke, James Szinger, and Hyejin Yoon

Project Officer: Stuart Shapiro, NIAID, NIH

Editors: Christian Apetrei, Beatrice Hahn, Ilene Mizrachi, James Mullins, Andrew Rambaut, Steve Wolinsky, Dan Barouch, Christian Brander, Rob De Boer, Bart Haynes, Richard Koup, John Moore, Bruce Walker, David Watkins

Theoretical Biology and Biophysics, T-6
Los Alamos National Laboratory

Workshop Topics

HIV Sequence Database and Immunology Database

Brian Foley and Karina Yusim

Session 1

Monday,
February 11
11:15 – 12:45

General introduction
Sequence search interface – alignments and basic trees
Geography search interface
Histogram
Database Alignments

Tools:

- *Genecutter - processing nucleotide sequences*
- *Neighbor Joining Treemaker*
- *HIV/SIV sequence locator tool*
- *Hypermut*
- *N-Glycosite*
- *Highlighter*
- *Protein Feature Accent*
- *Quality Control (if time permits)*

Los Alamos
NATIONAL LABORATORY

Workshop Topics

HIV Sequence Database and Immunology Database
Karina Yusim and Brian Foley

Session 2

Tuesday,
February 12
11:00 – 12:30

Immunology database introduction
Epitope maps and epitope summary tables
T-cell epitope search
T-cell epitope variants
Antibody search
List of most broadly neutralizing antibodies
HIV/SIV sequence locator tool
QuickAlign – Align an epitope to the database alignments
Motifscan – find HLA anchor residues in a protein
N-glycosite – finds N-linked glycosylation sites
ELF – epitope location finder
Peptgen – list peptides for reagent development

Mosaic Vaccine Maker, Epicover, and Posicover
- generate candidate vaccines
- estimate epitope coverage
- determine regional epitope coverage

Presenters:

Brian Foley: responsible for HIV Sequence Database and Vaccine Trials Database content and research; has a background in bioinformatics, lentivirus evolution and virology.

Karina Yusim: Co-PI of the project; Primary Editor for the HIV Immunology Database, responsible for HIV Immunology Database website and tools; has a background in sequence analysis, bioinformatics, immunology and applied mathematics.

seq-info@lanl.gov

Los Alamos
NATIONAL LABORATORY

Workshop Goals

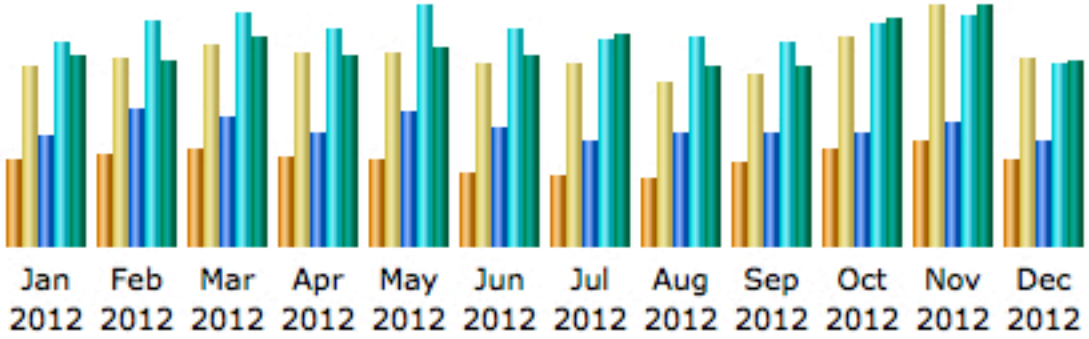
- Understanding the database content, how information was obtained, and what is available
- Database searching
- Examples of tools for analyses
- Quality control tools

The HIV Databases

- HIV Sequence database – founded 1986, G. Myers
  - Relational database, data from GenBank with added fields from the literature
  - Alignments – align indels and reduce multiple sequences per person
  - Annual hard copy and reviews
  - Web search interfaces: subtype, phenotype, geographic, sampling year…
  - Analysis tools
- HIV Immunology database – founded 1995, B. Korber
  - Comprehensive HIV epitope database, > 300-400 new papers per year
  - Integrate HIV immunological and sequence data
  - Annual hard copy and reviews
  - Web search interfaces: epitope, protein, HLA type, immunogen, keywords
  - Analysis tools for immunologists
- HIV Vaccine database – founded 2003, J. Mokili
  - A searchable relational database of published primate vaccine trials

Los Alamos
NATIONAL LABORATORY

Statistics: Monthly use

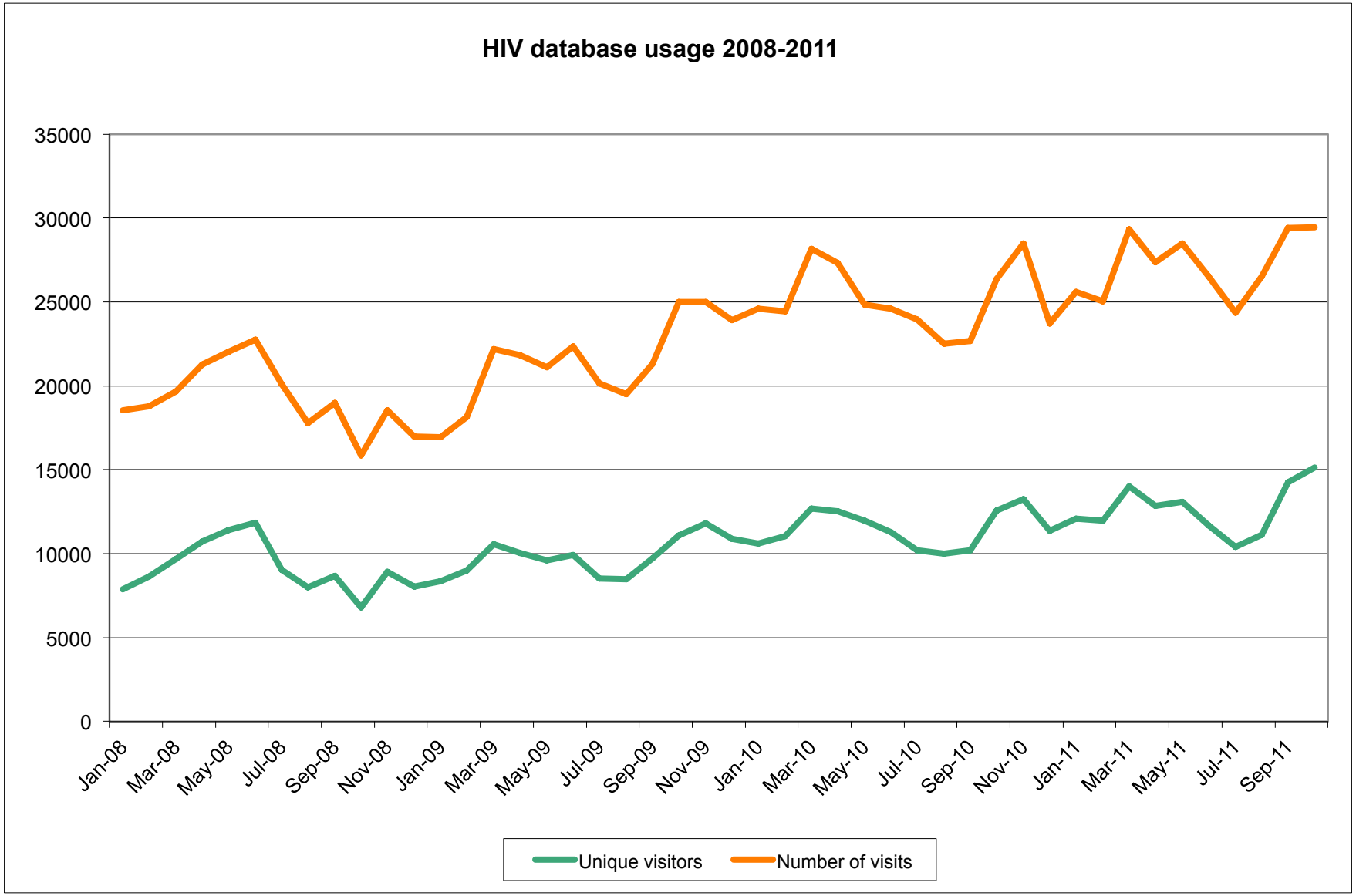


Month	Unique visitors	Number of visits	Pages	Hits	Bandwidth
Jan 2012	13,586	28,565	513,266	943,435	42.95 GB
Feb 2012	14,517	29,957	627,580	1,036,536	41.65 GB
Mar 2012	15,375	31,869	596,646	1,075,348	47.17 GB
Apr 2012	14,044	30,496	527,341	1,008,226	43.15 GB
May 2012	13,851	30,782	616,521	1,107,377	44.60 GB
Jun 2012	11,755	29,100	548,062	1,001,697	42.75 GB
Jul 2012	11,246	29,092	487,030	956,781	47.45 GB
Aug 2012	10,999	25,949	521,268	970,956	40.38 GB
Sep 2012	13,191	27,210	518,536	946,429	40.38 GB
Oct 2012	15,591	33,305	520,103	1,022,571	51.01 GB
Nov 2012	16,714	38,115	576,380	1,059,734	53.98 GB
Dec 2012	13,880	29,967	486,279	845,290	41.98 GB
Total	164,749	364,407	6,539,012	11,974,380	537.46 GB

http://www.hiv.lanl.gov/awstats/awstats.pl

Los Alamos NATIONAL LABORATORY

Database usage over time



http://www.hiv.lanl.gov/awstats/awstats.pl

Los Alamos NATIONAL LABORATORY

HIV DATABASES

Entry page at http://www.hiv.lanl.gov/

The **HIV databases** contain data on HIV genetic sequences, immunological epitopes, drug resistance-associated mutations, and vaccine trials. The website also gives access to a large number of tools that can be used to analyze these data. This project is funded by the Division of AIDS of the National Institute of Allergy and Infectious Diseases (NIAID), a part of the National Institutes of Health (NIH). Click on any of the links below to access a database. Editorial Board

SEQUENCE DATABASE ▸ | VACCINE DATABASE ▸

IMMUNOLOGY DATABASE ▸ | OTHER VIRUSES ▸

News: Archived News ▸

New Features for Epitope Location Finder (ELF)
ELF displays known and predicted epitopes found within a protein sequence query. ELF results now include both Class I (CTL) and Class II (helper) epitopes. In addition to predicting epitopes based on anchor residues, ELF now includes predictions from the Class I and Class II Binding Predictions tools at the Immune Epitope Database (IEDB). *13 March 2012*

New Features for HIV BLAST
HIV BLAST has new features. It now allows the user to find best matches among only subtyped sequences, or sequences of a specific subtype. It allows the resulting sequences to be downloaded fully aligned. *01 March 2012*

New Option for N-GlycoSite
The N-GlycoSite tool predicts N-linked glycosylation sites in amino acid sequences. A new option allows the user to exclude sites with a second-position proline, which is disfavored for N-linked glycosylation. *29 February 2012*

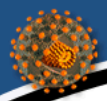
HIV Antibody Search Results More Specific
The antibody search interface in the HIV Immunology database is now more specific. Searches from the Author, Keyword, and Note fields now display only those notes and references that relate directly to the search. The user may still opt to display all, if desired *09 February 2012*

New Options for Quickalign
The Quickalign tool aligns any short protein or nucleotide sequence with database sequences. New options provide additional ways to calculate and display frequency by position, and allow the user to include the surrounding region in the alignment. *08 February 2012*

Questions or comments? Contact us at seq-info@lanl.gov

Los Alamos NATIONAL LABORATORY

Los Alamos NATIONAL LABORATORY



HIV sequence database

DATABASES | SEARCH | ALIGNMENTS | TOOLS | PUBLICATIONS | GUIDES | Search Site

- Search DB
- Advanced Search
- Intra-patient Search
- Next-gen Sequences
- Geography

HIV Sequence Database

Programs and Tools

Search Interface retrieves HIV and SIV sequences, which can then be aligned and used to build trees

Geography Search Interface retrieves HIV sequences based on geographical distribution

Tools for working with sequences lists all our online tools, organized by function

Alignments

HIV Premade Alignments includes Consensus and Ancestral Sequences, Subtype Reference Alignments, and Complete Alignments

Information

HIV Sequence Compendium print or order our annual publication

Tutorials and other information unpublished web-based content

Links to other HIV/AIDS tools and information

About this website

FAQ general information about this website

Site Statistics usage information for www.hiv.lanl.gov

How to Cite this Database

Editorial Board

News:

Archived News

New Features for HIV BLAST

HIV BLAST has new features. It now allows the user to find best matches among only subtyped sequences, or sequences of a specific subtype. It allows the resulting sequences to be downloaded fully aligned. *01 March 2012*

New Option for N-GlycoSite

The N-GlycoSite tool predicts N-linked glycosylation sites in amino acid sequences. A new option allows the user to exclude sites with a second-position proline, which is disfavored for N-linked glycosylation. *29 February 2012*

HIV Antibody Search Results More Specific

The antibody search interface in the HIV Immunology database is now more specific. Searches from the Author, Keyword, and Note fields now display only those notes and references that relate directly to the search. The user may still opt to display all, if desired *09 February 2012*

New Options for Quickalign

The Quickalign tool aligns any short protein or nucleotide sequence with database sequences. New options provide additional ways to calculate and display frequency by position, and allow the user to include the surrounding region in the alignment. *08 February 2012*

last modified: Tue Jan 26 10:10 2010

Questions or comments? Contact us at seq-info@lanl.gov.

DEPT OF HEALTH & HUMAN SERVICES

NATIONAL INSTITUTES OF HEALTH

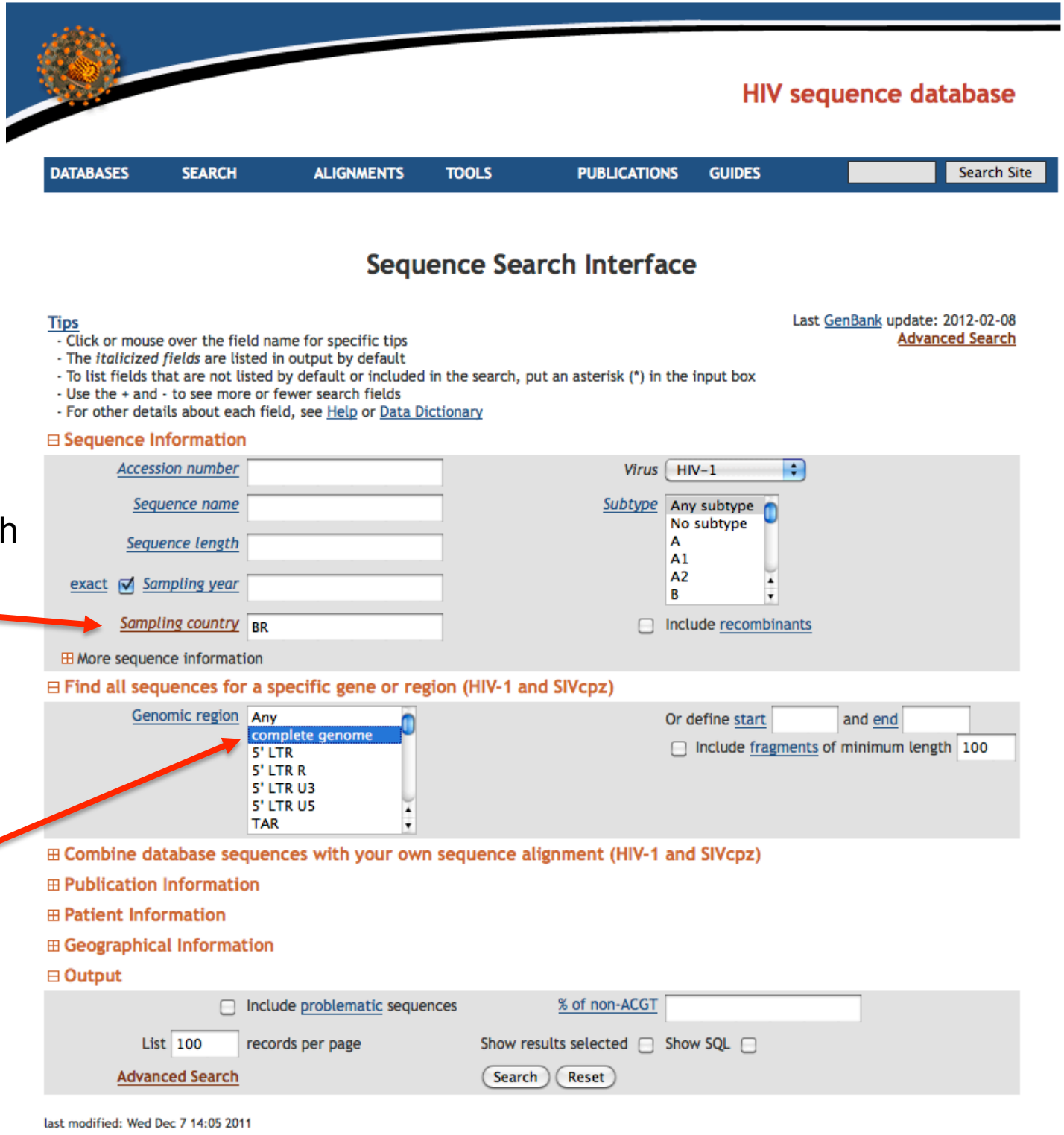
Search Interface

- **Help**
  - Tips at the top of the page are often overlooked
    - Ranges, operators, wildcards, logical groupings
  - Mouse-over provides brief descriptions; click field names for details in Help file
- **Searches**
  - Searches are case-insensitive
  - Records are searchable through sequence, patient, genomic region, or publication information and can be matched to the genomic region of a user-provided alignment
  - First seven fields will appear in search results page by default
  - A "*" in a textbox will cause that field to be included in the results page
  - Patient information (Infection year, Infection country) is different than sequence information (Sampling year and Sampling country)
  - Problematic sequence filters (hypermutation, frequent ambiguities, potential contamination)
- **Analysis**
  - Build a tree with user alignment, search results and subtype reference sequences combined
- **Results**
  - Can download aligned or unaligned sequences
  - Alignments are based on multiple pair wise alignments – alignments are good, but need hand editing for an optimal alignment
  - Select all or a subset of sequences for download
  - Sequences can be re-ordered by clicking on fields at the top of the page

Los Alamos
NATIONAL LABORATORY

We will search for country = Brazil (BR)

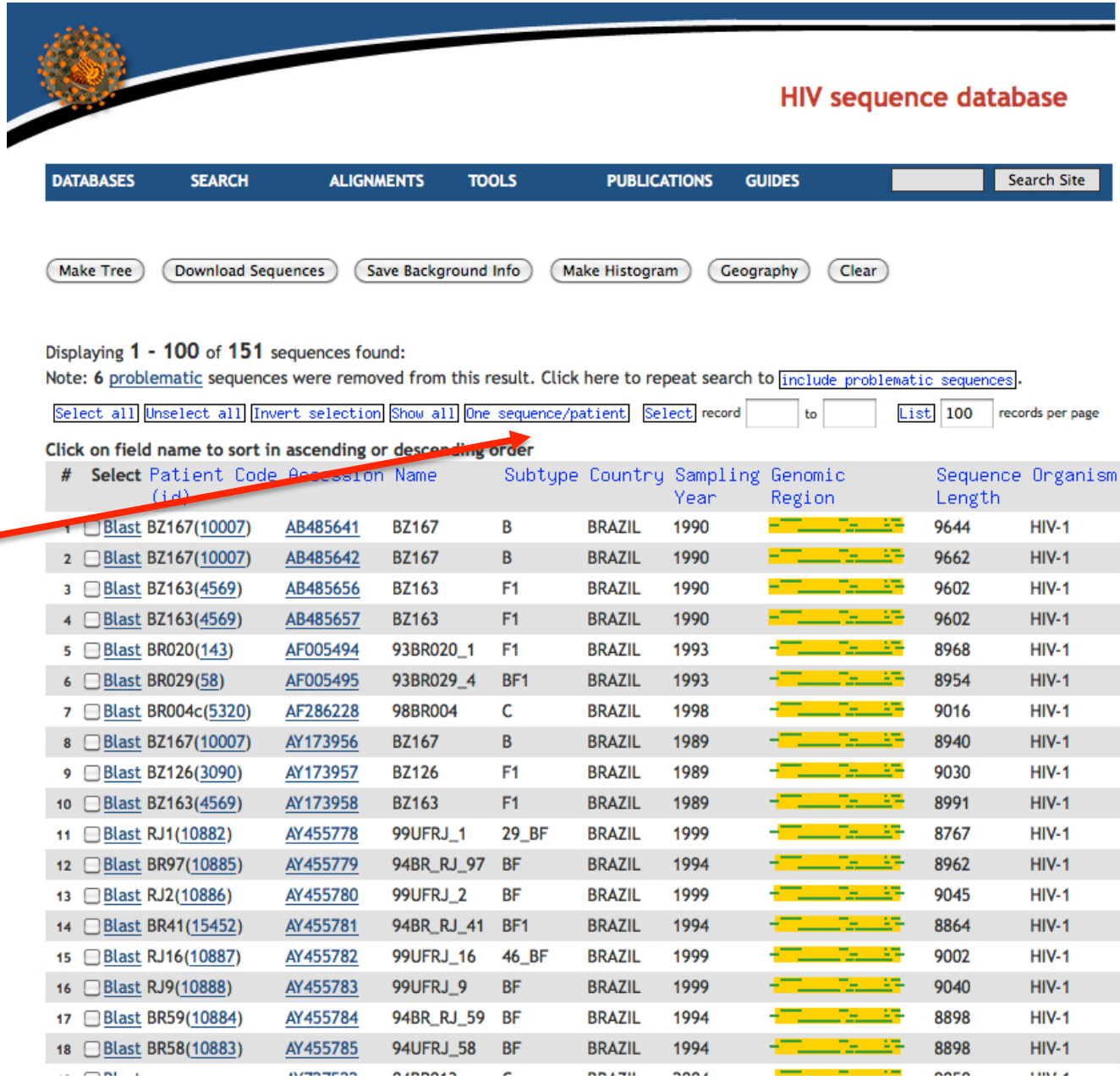
We will search for complete genomes.



Los Alamos NATIONAL LABORATORY

Questions or comments? Contact us at seq-info@lanl.gov.

Results for HIV-1 complete genomes from Brazil



HIV sequence database

DATABASES | SEARCH | ALIGNMENTS | TOOLS | PUBLICATIONS | GUIDES | Search Site

Make Tree | Download Sequences | Save Background Info | Make Histogram | Geography | Clear

Displaying **1 - 100** of **151** sequences found:

Note: **6** problematic sequences were removed from this result. Click here to repeat search to include problematic sequences.

Select all | Unselect all | Invert selection | Show all | One sequence/patient | Select record to List 100 records per page

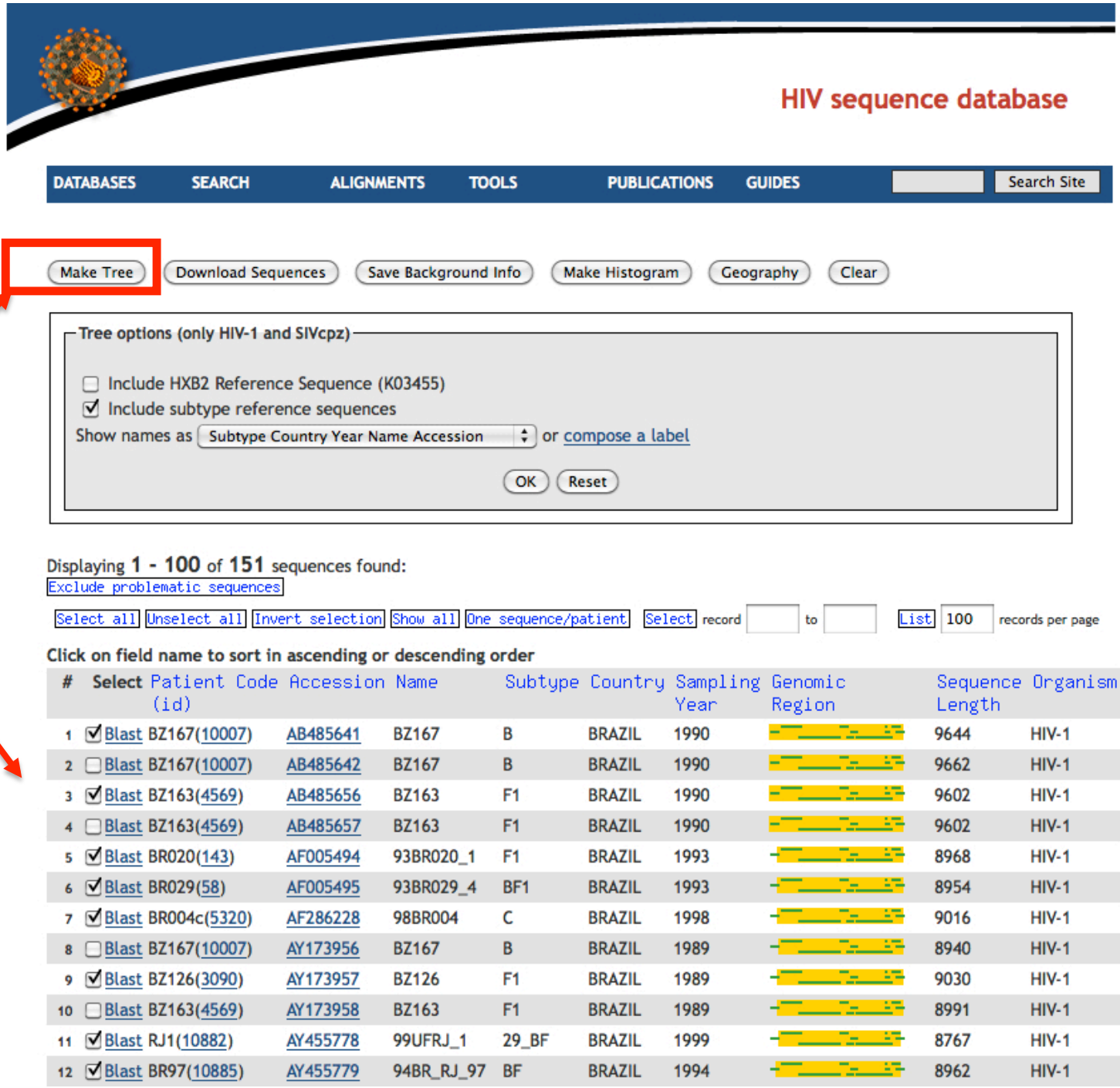
Click on field name to sort in ascending or descending order

#	Select	Patient Code (id)	Accession	Name	Subtype	Country	Sampling Year	Genomic Region	Sequence Length	Organism
1	☐	Blast BZ167(10007)	AB485641	BZ167	B	BRAZIL	1990		9644	HIV-1
2	☐	Blast BZ167(10007)	AB485642	BZ167	B	BRAZIL	1990		9662	HIV-1
3	☐	Blast BZ163(4569)	AB485656	BZ163	F1	BRAZIL	1990		9602	HIV-1
4	☐	Blast BZ163(4569)	AB485657	BZ163	F1	BRAZIL	1990		9602	HIV-1
5	☐	Blast BR020(143)	AF005494	93BR020_1	F1	BRAZIL	1993		8968	HIV-1
6	☐	Blast BR029(58)	AF005495	93BR029_4	BF1	BRAZIL	1993		8954	HIV-1
7	☐	Blast BR004c(5320)	AF286228	98BR004	C	BRAZIL	1998		9016	HIV-1
8	☐	Blast BZ167(10007)	AY173956	BZ167	B	BRAZIL	1989		8940	HIV-1
9	☐	Blast BZ126(3090)	AY173957	BZ126	F1	BRAZIL	1989		9030	HIV-1
10	☐	Blast BZ163(4569)	AY173958	BZ163	F1	BRAZIL	1989		8991	HIV-1
11	☐	Blast RJ1(10882)	AY455778	99UFRJ_1	29_BF	BRAZIL	1999		8767	HIV-1
12	☐	Blast BR97(10885)	AY455779	94BR_RJ_97	BF	BRAZIL	1994		8962	HIV-1
13	☐	Blast RJ2(10886)	AY455780	99UFRJ_2	BF	BRAZIL	1999		9045	HIV-1
14	☐	Blast BR41(15452)	AY455781	94BR_RJ_41	BF1	BRAZIL	1994		8864	HIV-1
15	☐	Blast RJ16(10887)	AY455782	99UFRJ_16	46_BF	BRAZIL	1999		9002	HIV-1
16	☐	Blast RJ9(10888)	AY455783	99UFRJ_9	BF	BRAZIL	1999		9040	HIV-1
17	☐	Blast BR59(10884)	AY455784	94BR_RJ_59	BF	BRAZIL	1994		8898	HIV-1
18	☐	Blast BR58(10883)	AY455785	94UFRJ_58	BF	BRAZIL	1994		8898	HIV-1

Choose "One sequence/ patient" to remove very similar sequences (only available if a region is selected)

Los Alamos NATIONAL LABORATORY

Select a few sequences and make tree, allows us to add a reference set to our data and align them



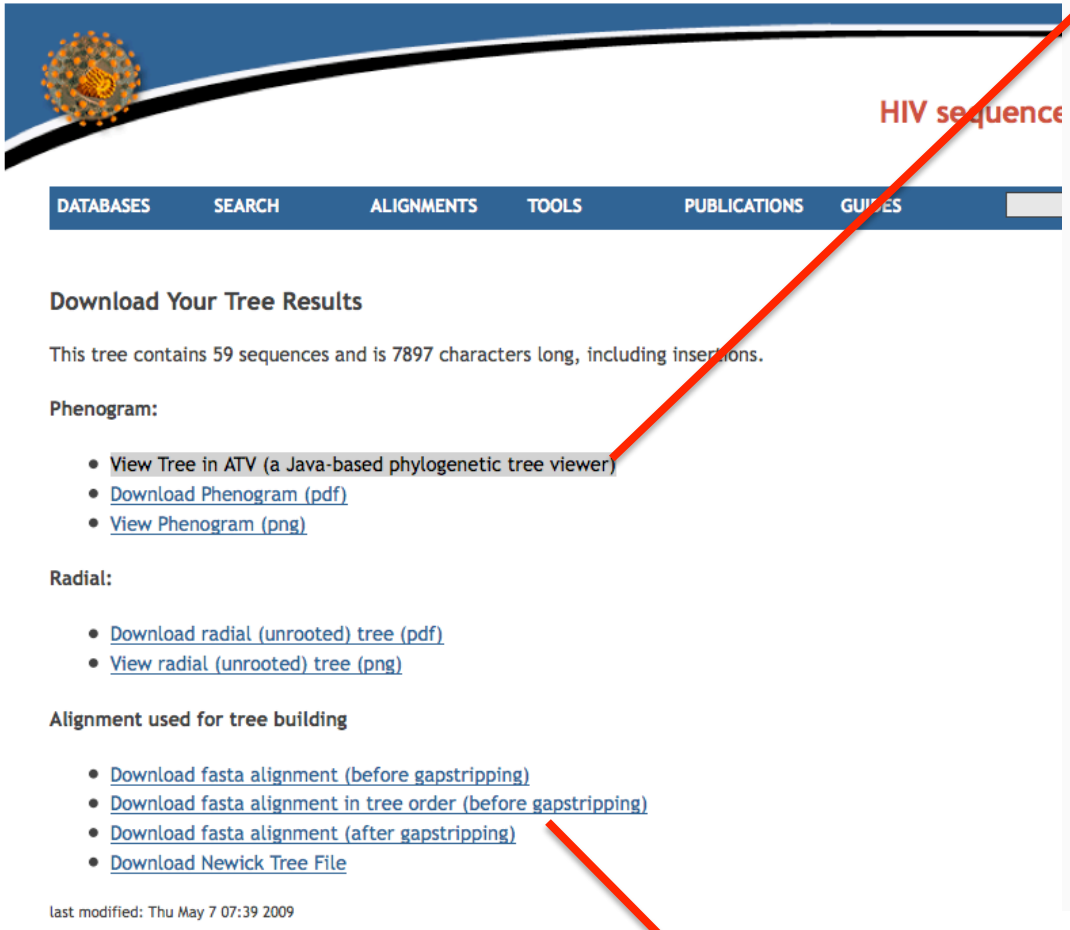
#	Select	Patient Code (id)	Accession	Name	Subtype	Country	Sampling Year	Genomic Region	Sequence Length	Organism
1	☑	Blast BZ167(10007)	AB485641	BZ167	B	BRAZIL	1990		9644	HIV-1
2	☐	Blast BZ167(10007)	AB485642	BZ167	B	BRAZIL	1990		9662	HIV-1
3	☑	Blast BZ163(4569)	AB485656	BZ163	F1	BRAZIL	1990		9602	HIV-1
4	☐	Blast BZ163(4569)	AB485657	BZ163	F1	BRAZIL	1990		9602	HIV-1
5	☑	Blast BR020(143)	AF005494	93BR020_1	F1	BRAZIL	1993		8968	HIV-1
6	☑	Blast BR029(58)	AF005495	93BR029_4	BF1	BRAZIL	1993		8954	HIV-1
7	☑	Blast BR004c(5320)	AF286228	98BR004	C	BRAZIL	1998		9016	HIV-1
8	☐	Blast BZ167(10007)	AY173956	BZ167	B	BRAZIL	1989		8940	HIV-1
9	☑	Blast BZ126(3090)	AY173957	BZ126	F1	BRAZIL	1989		9030	HIV-1
10	☐	Blast BZ163(4569)	AY173958	BZ163	F1	BRAZIL	1989		8991	HIV-1
11	☑	Blast RJ1(10882)	AY455778	99UFRJ_1	29_BF	BRAZIL	1999		8767	HIV-1
12	☑	Blast BR97(10885)	AY455779	94BR_RJ_97	BF	BRAZIL	1994		8962	HIV-1

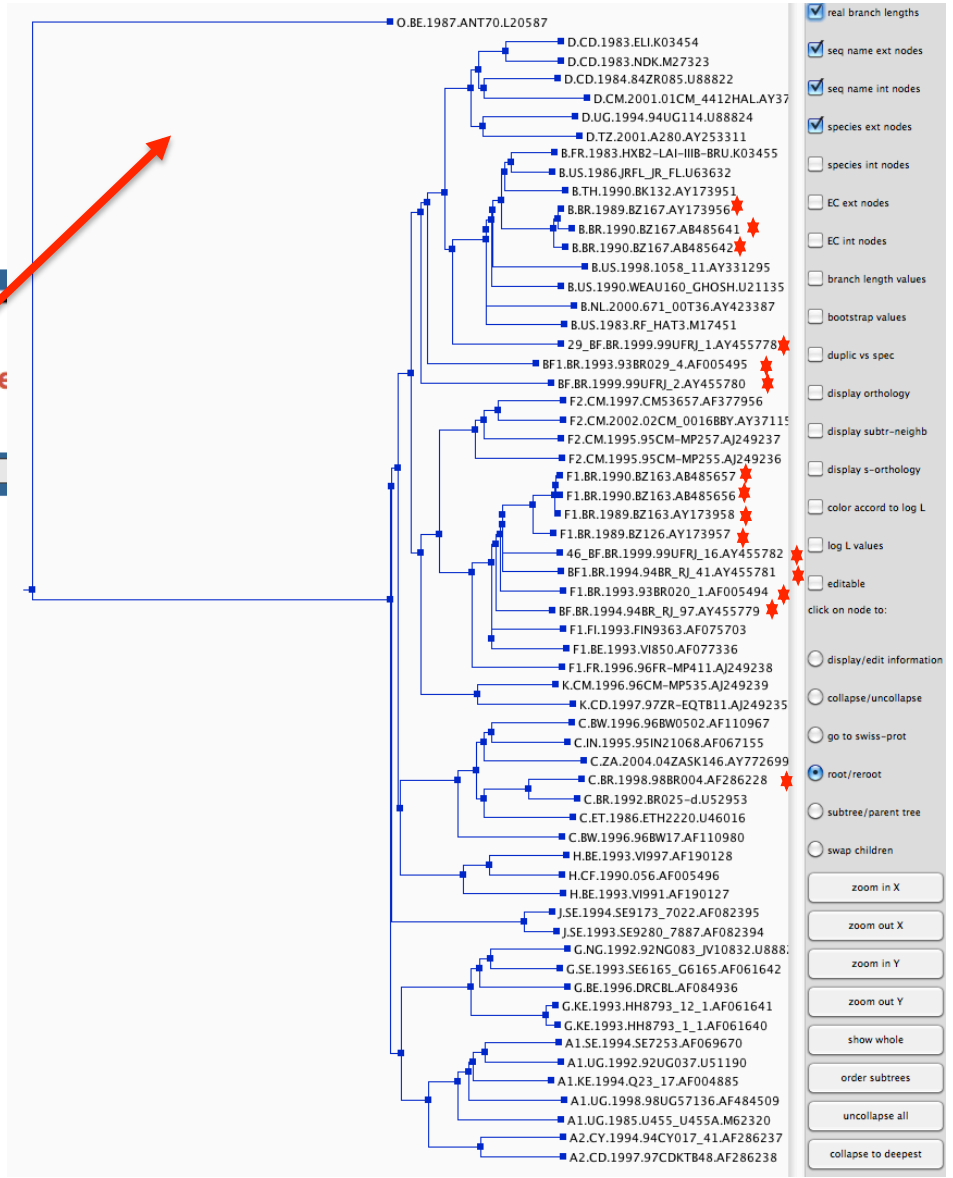
TreeMaker tool

Choice of outgroup influences the the tree. In general, choose next closest sequences to the “ingroup”. In this case our Brazilian sequences are all HIV-1 M group.

Optional mailback, and tree title

ATV java-based view for quick look, cannot save/print





Obtaining your sequences of interest and having them aligned to a good reference set was the whole point of this. The tree was just a first check on data and alignment quality.

Save alignment, use BioEdit or SeAl to view/adjust.

Los Alamos NATIONAL LABORATORY

HIV sequence database

DATABASES SEARCH ALIGNMENTS TOOLS PUBLICATIONS GUIDES Search Site

Download Your Tree Results

This tree contains 59 sequences and is 7897 characters long, including insertions.

Phenogram:

- View Tree in ATV (a Java-based phylogenetic tree viewer)
- Download Pheno
- View Phenogram

Radial:

- Download radial
- View radial (unr

Alignment used for tre

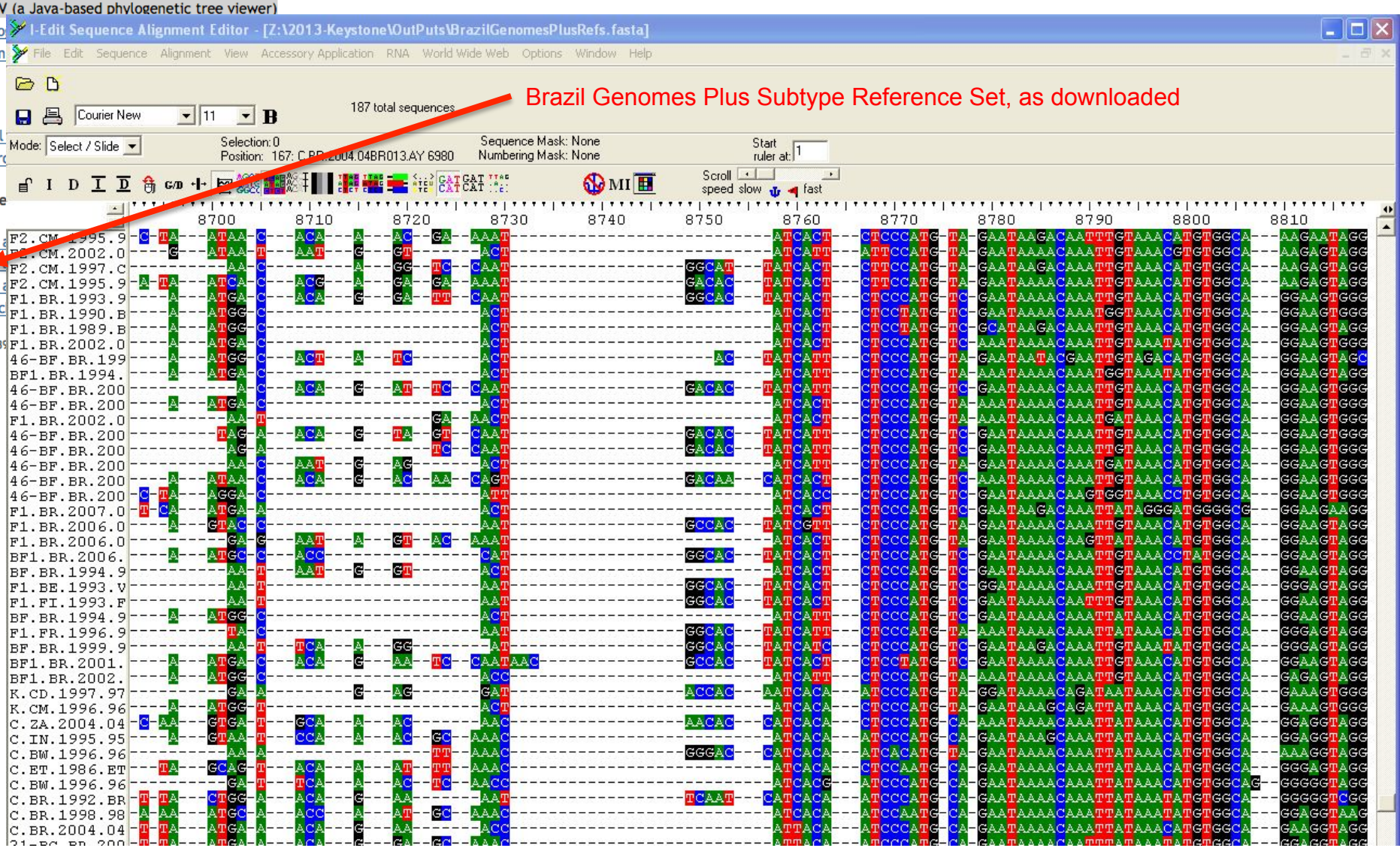
- Download fasta
- Download fast
- Download fasta
- Download Newic

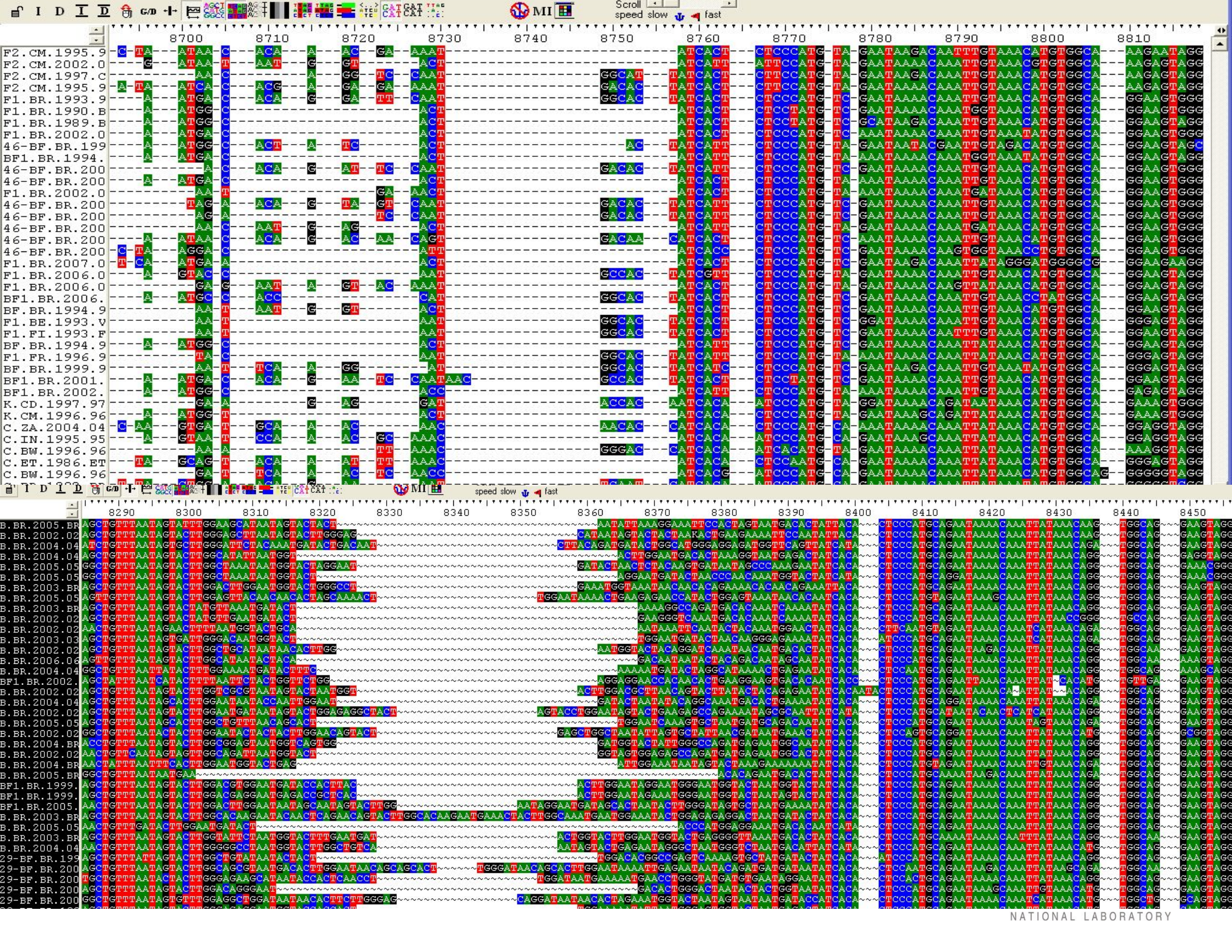
last modified: Thu May 7 07:39

Save alignment, use BioEdit or SeAl to view/adjust.

Send alignment to GeneCutter or HIV-Align first, is usually best.

http://www.hiv.lanl.gov/content/sequence/GENE_CUTTER/cutter.html




Scroll
speed slow
fast
MI
8700
8710
8720
8730
8740
8750
8760
8770
8780
8790
8800
8810
F2.CM.1995.9
F2.CM.2002.0
F2.CM.1997.C
F2.CM.1995.9
F1.BR.1993.9
F1.BR.1990.B
F1.BR.1989.B
F1.BR.2002.0
46-BF.BR.199
BF1.BR.1994.
46-BF.BR.200
46-BF.BR.200
F1.BR.2002.0
46-BF.BR.200
46-BF.BR.200
46-BF.BR.200
46-BF.BR.200
46-BF.BR.200
F1.BR.2007.0
F1.BR.2006.0
F1.BR.2006.0
BF1.BR.2006.
BF.BR.1994.9
F1.BE.1993.V
F1.FI.1993.F
BF.BR.1994.9
F1.FR.1996.9
BF.BR.1999.9
BF1.BR.2001.
BF1.BR.2002.
K.CD.1997.97
K.CM.1996.96
C.ZA.2004.04
C.IN.1995.95
C.BW.1996.96
C.ET.1986.ET
C.BW.1996.96
speed slow
fast
MI
8290
8300
8310
8320
8330
8340
8350
8360
8370
8380
8390
8400
8410
8420
8430
8440
8450
B.BR.2005.BR
B.BR.2002.02
B.BR.2004.04
B.BR.2004.04
B.BR.2005.05
B.BR.2005.05
B.BR.2003.BR
B.BR.2005.05
B.BR.2003.BR
B.BR.2002.02
B.BR.2002.02
B.BR.2003.03
B.BR.2002.02
B.BR.2006.06
B.BR.2004.04
BF1.BR.2002.
B.BR.2002.02
B.BR.2004.04
B.BR.2002.02
B.BR.2005.05
B.BR.2002.02
B.BR.2004.BR
B.BR.2002.02
B.BR.2004.BR
B.BR.2005.BR
BF1.BR.1999.
BF1.BR.1999.
BF1.BR.2005.
B.BR.2003.BR
B.BR.2005.05
B.BR.2003.BR
B.BR.2004.04
29-BF.BR.199
29-BF.BR.200
29-BF.BR.200
29-BF.BR.200
29-BF.BR.200
NATIONAL LABORATORY

New search: all complete genomes; then look at geographic and subtype distribution of the sequences

Displaying **1 - 100** of **5338** sequences found:

Note: **478** problematic sequences were removed from this result. Click here to repeat search to include problematic sequences.

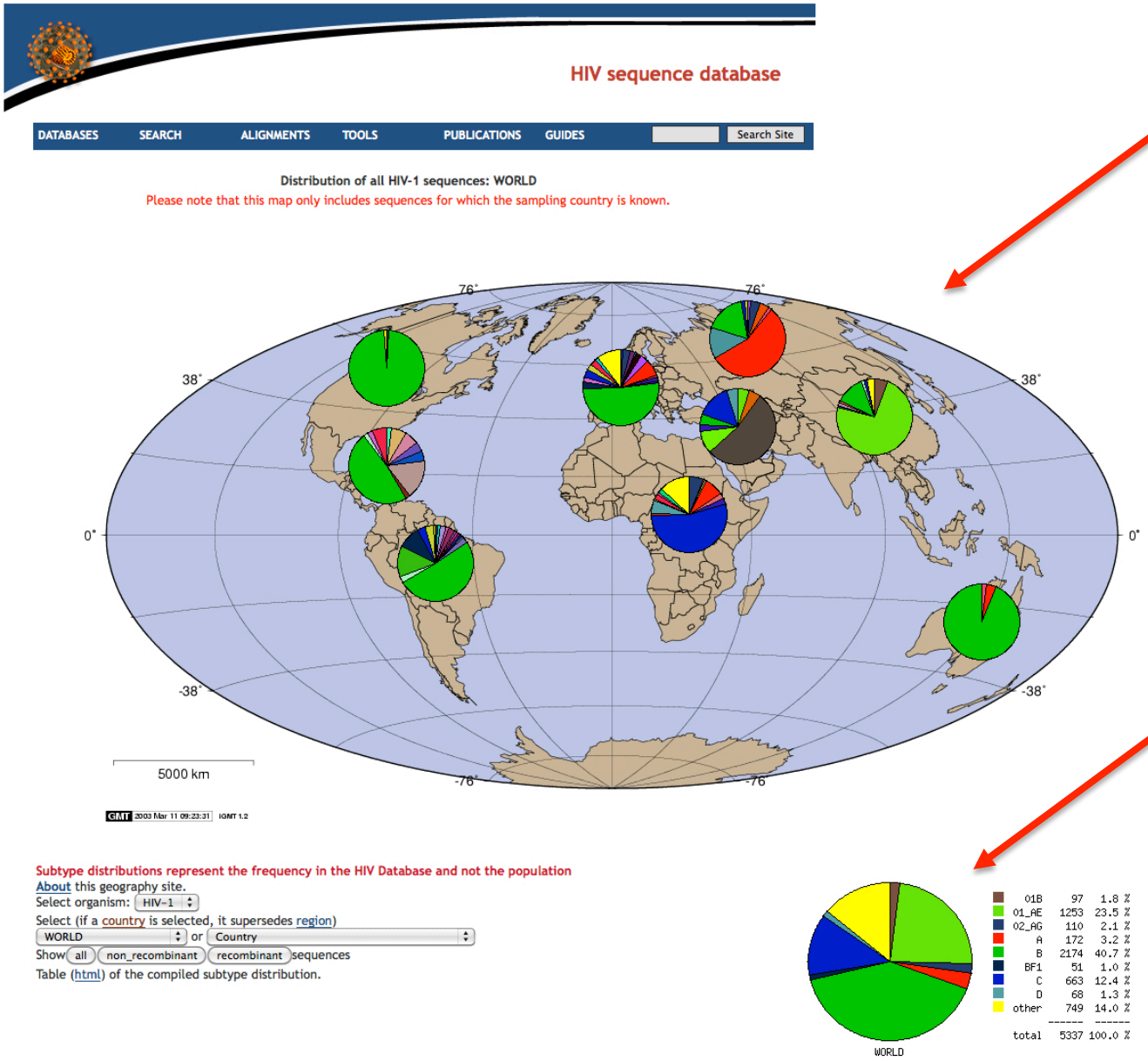
Select all | Unselect all | Invert selection | Show all | Select record to | List 100 records per page

Click on field name to sort in ascending or descending order

#	Select	Patient Code (id)	Accession	Name	Subtype	Country	Sampling Year	Genomic Region	Sequence Length	Organism
1	Blast	LAI(19535)	A04321	IIIB_LAI	B	FRANCE	1983		9193	HIV-1
2	Blast	ELI(580)	A07108	ELI_patent	D	DEM REP OF CONGO	1983		9176	HIV-1
3	Blast	MAL(578)	A07116	MAL_patent	A1DK	DEM REP OF CONGO	1985		9229	HIV-1
4	Blast	LAI(19535)	A07867	LAI-J19	B	FRANCE	1983		9193	HIV-1
5	Blast	ELI(580)	A14116	ELI_patent	D	DEM REP OF CONGO	1983		9176	HIV-1
6	Blast	NDK(13796)	A34828	NDK_patent	D	DEM REP OF CONGO	1983		9143	HIV-1
7	Blast	IN101(14294)	AB023804	93IN101	C	INDIA	1993		9680	HIV-1
8	Blast	C1_husband(15892)	AB032740	95TNIH022	01_AE	THAILAND	1995		9427	HIV-1
9	Blast	47(881)	AB032741	95TNIH047	01_AE	THAILAND	1995		9430	HIV-1
10	Blast	NJ97-42(24045)	AB049811	97GH-AG1	02_AG	GHANA	1997		9748	HIV-1
11	Blast		AB052867	97GH-AG2	02A1	GHANA	1997		9708	HIV-1
12	Blast	NH1(717)	AB052995	93JP_NH1	01_AE	JAPAN	1993		9720	HIV-1
13	Blast	NH2(715)	AB070352	NH25_93JPNH25T_93JP_NH2_5T	01_AE	JAPAN	1993		9731	HIV-1
14	Blast	CS2(9760)	AB078005	ARES2	B	UNITED STATES	1997		9637	HIV-1
15	Blast	502(3272)	AB097865	mIDU502	01B	MYANMAR	2000		9046	HIV-1

Los Alamos NATIONAL LABORATORY

Geography output



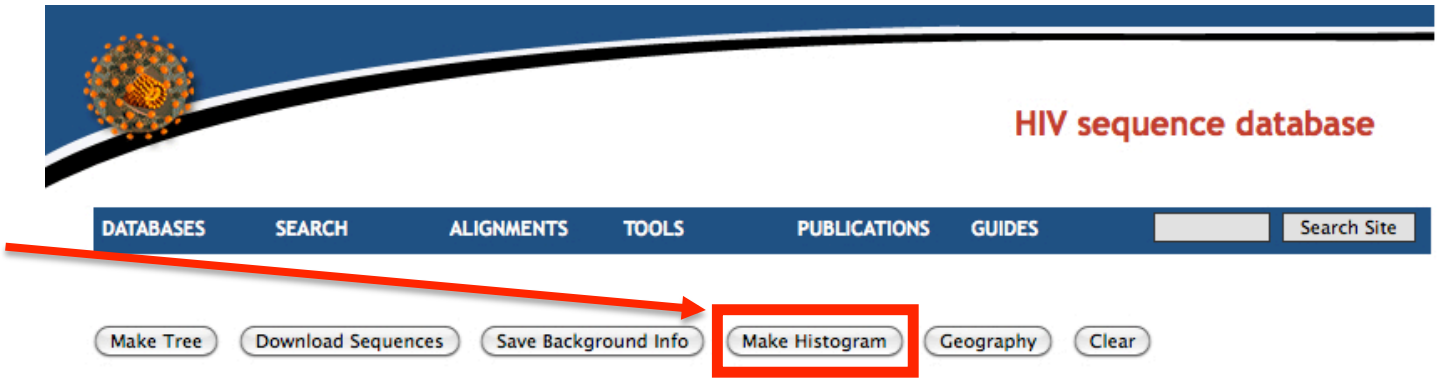
Each continent's pie chart is clickable to "zoom in" on that continent.

Likewise for each country once you are zoomed in to the continent level.

Most complete genomes in the HIV database are subtype B. But subtype C is more prevalent in human infections. Beware of this type of sampling bias.

Los Alamos NATIONAL LABORATORY

New search: all sequences from Brazil. Then look at the distribution of the sequences over the genome



Displaying **1 - 100** of **15489** sequences found:

Note: **87** problematic sequences were removed from this result. Click here to repeat search to include problematic sequences.

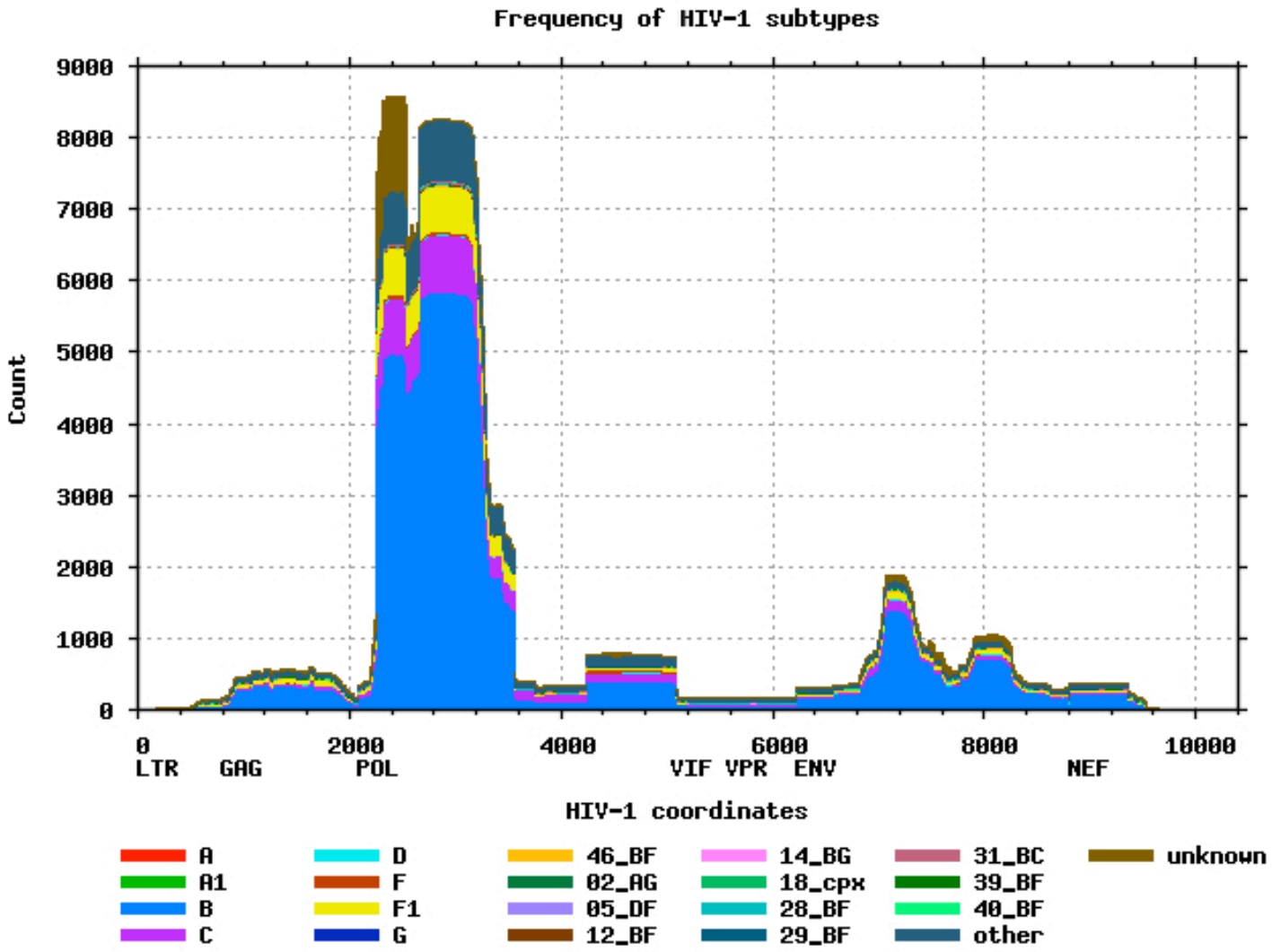
Select all | Unselect all | Invert selection | Show all | Select record to | List 100 records per page

Click on field name to sort in ascending or descending order

#	Select	Patient Code (id)	Accession	Name	Subtype	Country	Sampling Year	Genomic Region	Sequence Length	Organism
1	Blast	BZ167(10007)	AB485641	BZ167	B	BRAZIL	1990		9644	HIV-1
2	Blast	BZ167(10007)	AB485642	BZ167	B	BRAZIL	1990		9662	HIV-1
3	Blast	BZ163(4569)	AB485656	BZ163	F1	BRAZIL	1990		9602	HIV-1
4	Blast	BZ163(4569)	AB485657	BZ163	F1	BRAZIL	1990		9602	HIV-1
5	Blast	RJ100(4)	AF000238	RJ100	D	BRAZIL	1996		424	HIV-1
6	Blast	BR020(143)	AF005494	93BR020_1	F1	BRAZIL	1993		8968	HIV-1
7	Blast	BR029(58)	AF005495	93BR029_4	BF1	BRAZIL	1993		8954	HIV-1
8	Blast	BR003(655)	AF009369	92BR003	B	BRAZIL	1992		1176	HIV-1
9	Blast	BR004a(656)	AF009370	92BR004	B	BRAZIL	1992		1175	HIV-1
10	Blast	BR017(657)	AF009371	92BR017_A	B	BRAZIL	1992		1174	HIV-1
11	Blast	BR018(658)	AF009372	92BR018_A	B	BRAZIL	1992		1174	HIV-1
12	Blast	92BR019(72)	AF009373	92BR019_A	B	BRAZIL	1992		1176	HIV-1
13	Blast	92BR020(8574)	AF009374	92BR020_A	B	BRAZIL	1992		1176	HIV-1
14	Blast	BR021(8563)	AF009375	92BR021a	B	BRAZIL	1992		1173	HIV-1

Los Alamos NATIONAL LABORATORY

Histogram output



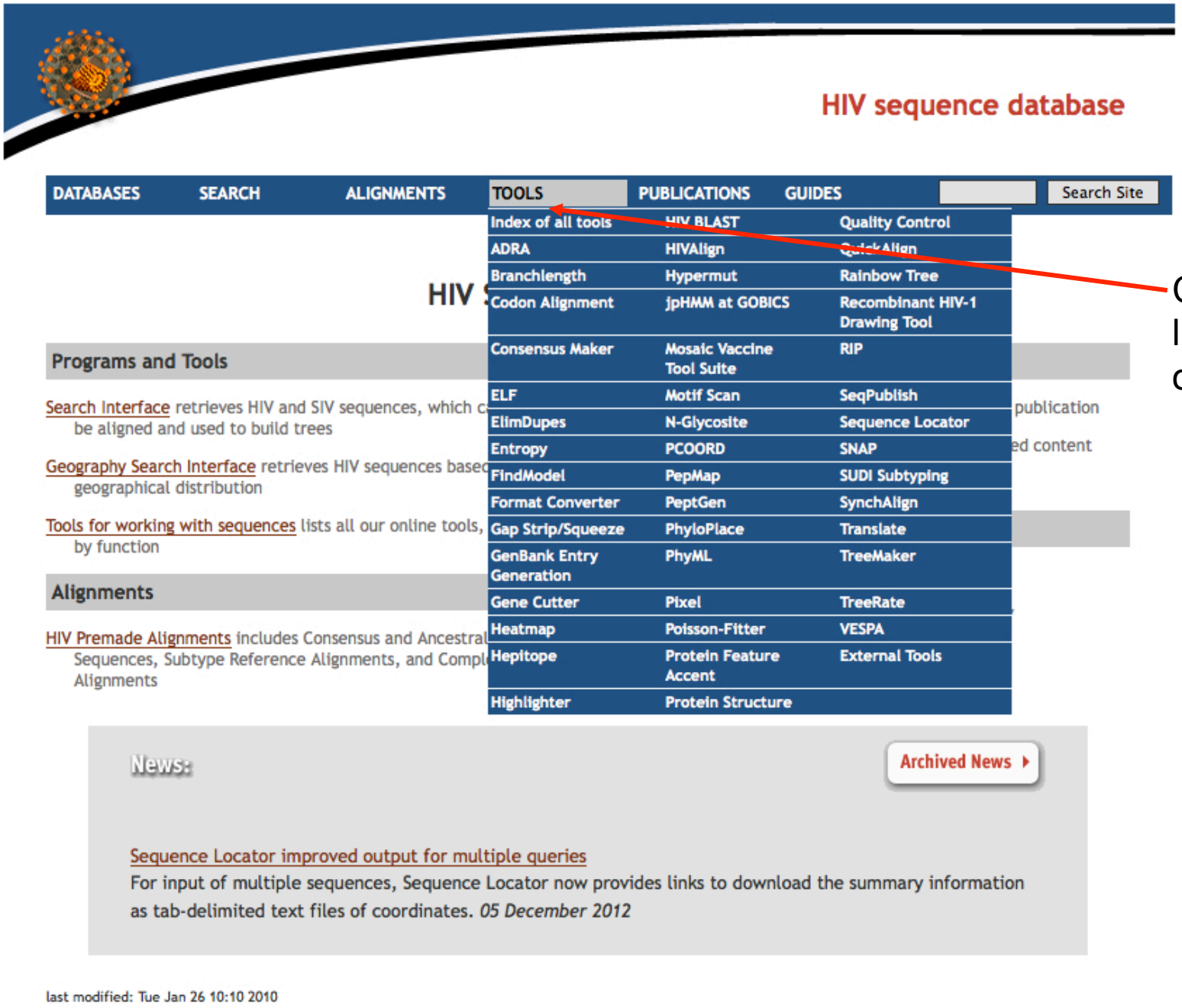
This histogram shows the distribution of sequences from your query across the entire HIV-1 genome. At each position across the genome, the number of sequences overlapping with that position is plotted. The colors represent different subtypes.

Los Alamos
NATIONAL LABORATORY

Tools

- **Analysis and Quality Control**
  - **HIV BLAST** finds sequences similar to yours in the HIV database.
  - **N-Glycosite** finds potential N-linked glycosylation sites.
  - **RIP 3.0** (Recombinant Identification Program) detects HIV-1 subtypes and recombination.
- **Alignment and sequence manipulation**
  - **HIValign** uses our HMM alignment models to align your sequences.
  - **Gapstreeze** removes columns with more than a given % of gaps.
  - **EpimDupes** Given an alignment or set of unaligned nucleotide or protein sequences, this tool compares the sequences and eliminates any duplicates.
- **Phylogenetics**
  - **TreeMaker** generates a neighbor-joining phylogenetic tree.
  - **PhyML** generates a maximum likelihood phylogenetic tree.
  - **TreeRate** finds the phylogenetic root of a tree and calculates evolutionary rate.
- **Format and display**
  - **Protein Feature Accent** provides an interactive 3-D graphic of HIV proteins; the user can map a sequence feature (a short functional domain, epitope, or amino acid) and see where it occurs spatially in the 3D structure.
  - **Highlighter** highlights mismatches, matches, transition and transversion mutations, and silent and non-silent mutations in an alignment of nucleotide sequences.
  - **SeqPublish** makes alignment publication-ready.
  - **Recombinant HIV drawing tool** highlights regions of the genome on a graphically representation

The HIV database sequence analysis tool set



Click top level to link to full page of tools

Los Alamos NATIONAL LABORATORY

HIV Database Tools

(alphabetical order within category)

For detailed descriptions, mouse over the links.

Analysis and Quality Control

Entropy quantifies positional variation in an alignment using Shannon Entropy

HIV BLAST finds sequences similar to yours in the HIV database

Hypermut detects hypermutation

jpHMM at GOBICS detects subtype recombination in HIV-1; hosted at GOBICS as a collaboration between the Department of Bioinformatics, University of Göttingen and the Los Alamos HIV Sequence Database

N-Glycosite finds potential N-linked glycosylation sites

PCOORD multidimensional analysis of sequence variation

Quality Control runs several tools to allow quick QC analysis of HIV-1 sequences; optional step prepares sequence submission for GenBank

RIP (Recombinant Identification Program) detects HIV-1 subtypes and recombination

SNAP calculates synonymous/non-synonymous substitution rates

SUDI Subtyping plots the distance of your sequence to established subtypes

VESPA (Viral Epidemiology Signature Pattern Analysis) detects residues with different frequencies in two sequence sets

Alignment and sequence manipulation

Codon Alignment takes a nucleotide alignment and returns a codon alignment and translation

Consensus Maker computes a customizable consensus

ElimDupes compares the sequences within an alignment and eliminates any duplicates

Gap Strip/Squeeze removes columns with more than a given % of gaps

Gene Cutter clips genes from a nucleotide alignment, codon-aligns, and translates

HIValign uses our HMM alignment models to align your sequences

Phylogenetics

Branchlength calculates branch lengths between internal and end nodes

FindModel finds which evolutionary model best fits your sequences

PhyloPlace reports phylogenetic relatedness of an HIV-1 sequence with reference sequences

PhyML generates much better trees than our simple TreeMaker tool

Poisson-Fitter estimates time since MRCA and star-phylogeny. For use with acute (low diversity) samples.

TreeMaker generates a quick-and-dirty phylogenetic tree

TreeRate finds the phylogenetic root of a tree and calculates evolutionary rate

Immunology

ELF (Epitope Location Finder) identifies known and potential epitopes within peptides

Epilign (QuickAlign) aligns a protein sequence (e.g., epitope) to the appropriate protein alignment

Heatmap displays a table of numbers by using colors to represent the numerical values

Hepitope identifies potential epitopes based on HLA frequencies

Mosaic Vaccine Tool Suite designs and assesses polyvalent protein sequences for T-cell vaccines

Motif Scan finds HLA anchor motifs in protein sequences for specified HLA serotypes, genotypes or supertypes

PeptGen generates overlapping peptides from a protein sequence

Database search interfaces

ADRA Antiviral Drug Resistance Analysis, a resistance mutation database

Advanced Search creates a custom search interface

Tools are organized in groups by function/purpose.

Most tools have explanation pages, and sample data sets.

Many tools were inspired by user comments, please ask for more.

Los Alamos
NATIONAL LABORATORY

SynchAlign aligns overlapping alignments to one another

QuickAlign (formerly Epilign and Primalign) aligns a nucleotide or protein sequence (e.g., primer or epitope) to the appropriate genome alignment

Codon Alignment takes a nucleotide alignment and returns a codon alignment and translation

ElimDupes compares the sequences within an alignment and eliminates any duplicates

Pixel generates a PNG image of an alignment using 1 or more colored pixel(s) for each residue

PepMap can be used to map epitopes, functional domains, or any protein region of interest

Format and display

Protein Feature Accent provides an interactive 3-D graphic of HIV proteins; can map a sequence feature (a short functional domain, epitope, or amino acid) and see it spatially

Format Converter converts between alignment formats

SeqPublish makes publication-ready alignments

Highlighter highlights mismatches, matches, transitions and transversion mutations and silent and non-silent mutations in an alignment of nucleotide sequences

Recombinant HIV-1 Drawing Tool creates a graphical representation of your HIV-1 intersubtype recombinant

Protein Structure Analysis provides a visualization tool for protein sequence properties

Advanced Search creates a custom search interface

Geography shows the geographic distribution of sequences in the database

CTL/CD8+ Search searches for CD8+ epitopes by protein, immunogen, HLA, author, keywords

T-Helper/CD4+ Search search for CD4+ epitopes by protein, immunogen, HLA, author, keywords

Antibodies search for HIV antibodies by protein, immunogen, AB type, isotype, author, keywords

Vaccine Trials Database finds past vaccine trials and their results

ADRA Antiviral Drug Resistance Analysis, a resistance mutation database

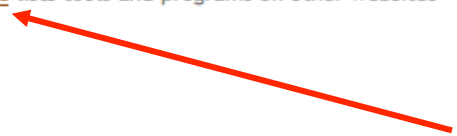
Other tools

HDent and HDdist perform analysis of heteroduplex mobility shifts

ODprep and ODfit calculate antibody titers based on concentration and optical density data

External tools

External tools lists tools and programs on other websites



We tend to list only tools of great use in HIV research. Many of these tools are essential, such as either BioEdit or SeAl for alignment viewing and correction.

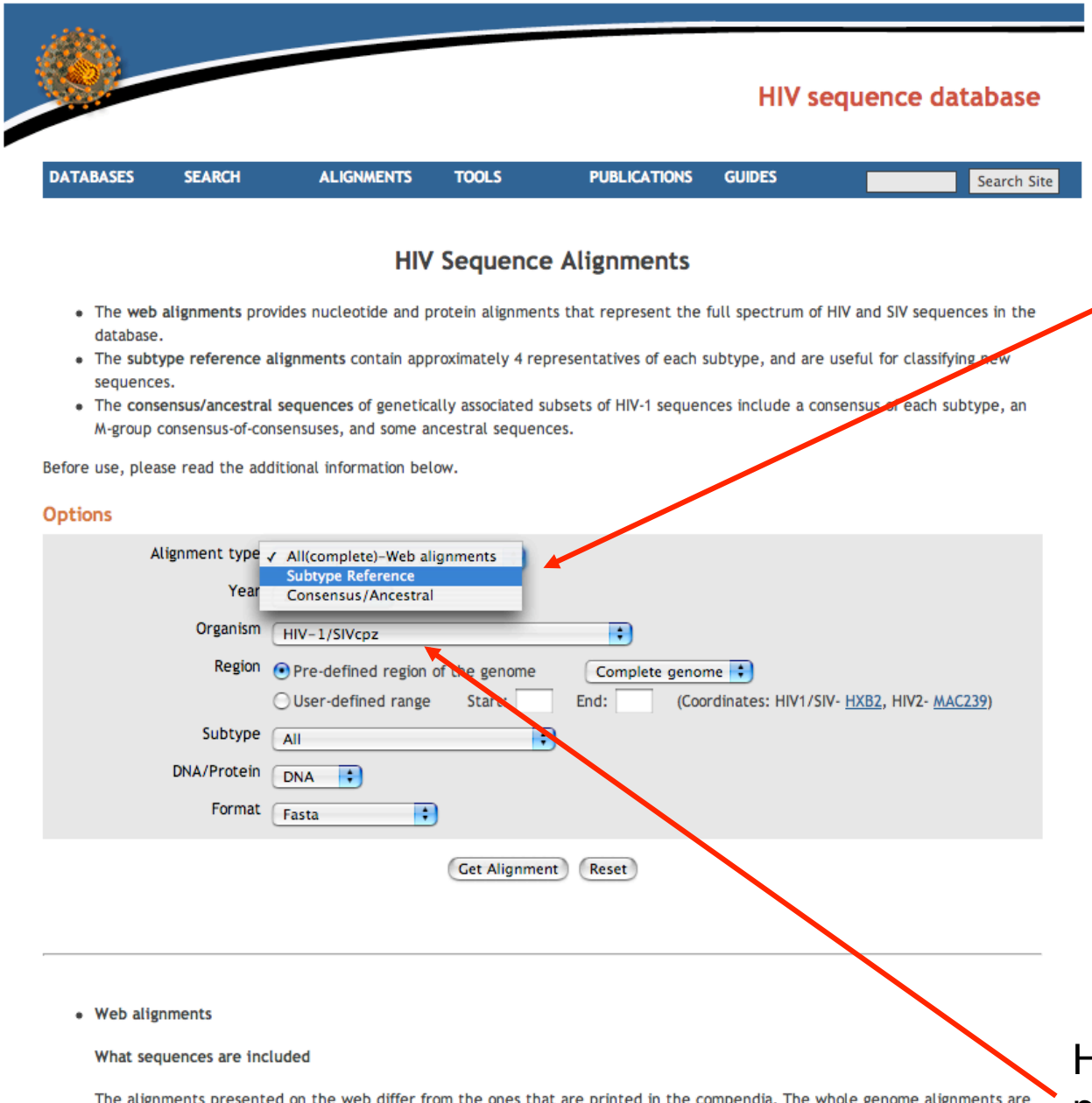
http://www.hiv.lanl.gov/content/sequence/HIV/HIVTools.html

Los Alamos NATIONAL LABORATORY

Pre-Built Sequence alignments

- Originally based on iterations of manual and HMM alignments
- Yearly updates using HMM and manual corrections
- Alignments are in reading frame (codon aligned)
- Contain non-redundant data (one sequence per patient)
- Compendium alignments show fewer sequences than web version
- Reference alignments contain up to four representatives of each subtype. One of each CRF.
- Protein alignments may contain frameshift compensations
- Subtype consensus with ties resolved, as well as maximum likelihood ancestors, are available for reagent production
- Special interest alignments are being added
  - Sequence sets of particular research interest
  - Suggestions welcome to tkl@lanl.gov

Los Alamos
NATIONAL LABORATORY



Alll(complete) = one per patient, all sequences for which we have a complete genome, or a complete gene.

Subtype Reference = 4 representatives of each subtype, plus one of each Circulating intersubtype recombinant form (CRF) of the M group, plus 4 O group, N group, P group and SIV-CPZ

Consensus/Ancestral computed from master alignment periodically.

HIV-2/SIV-SMM and primate lentivirus alignments also available here.

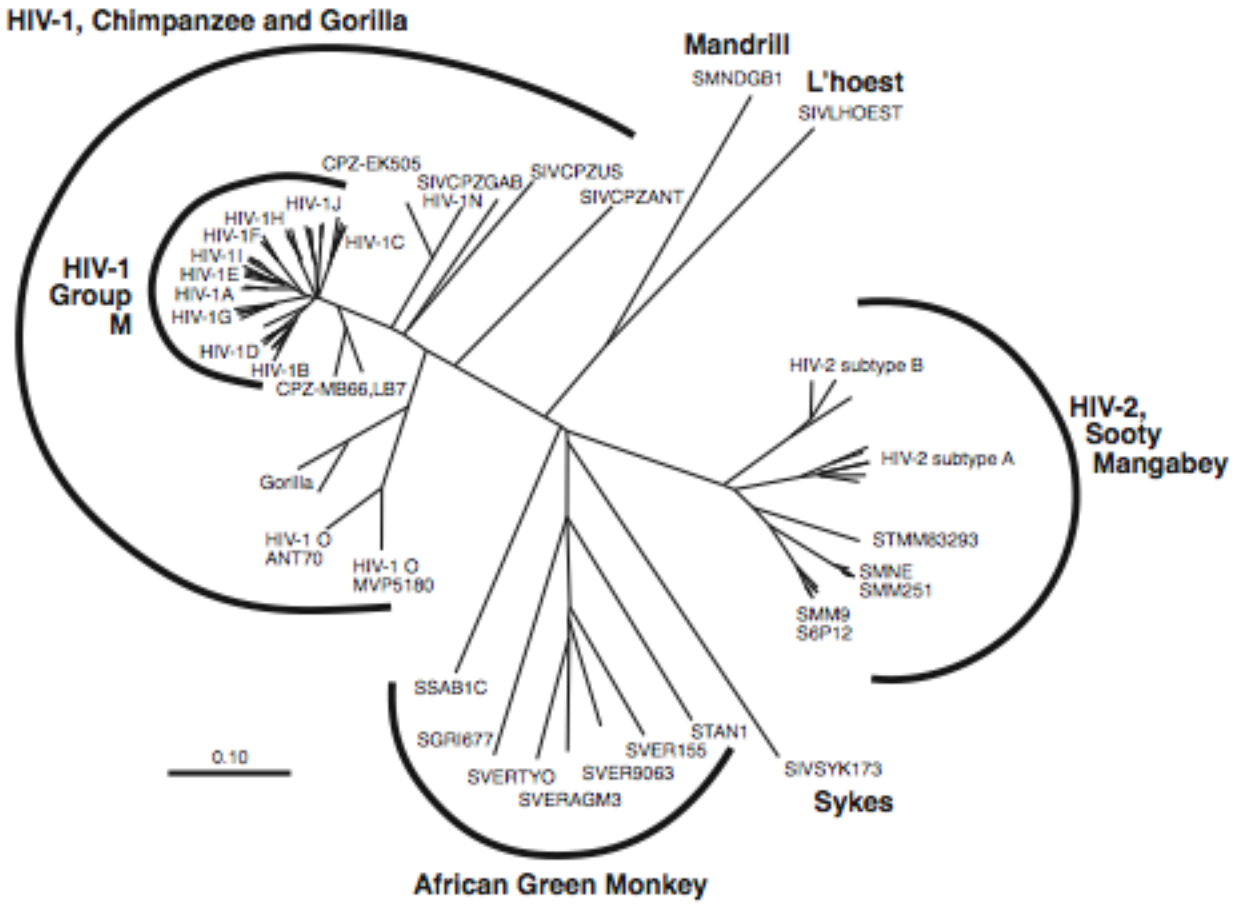
Los Alamos NATIONAL LABORATORY

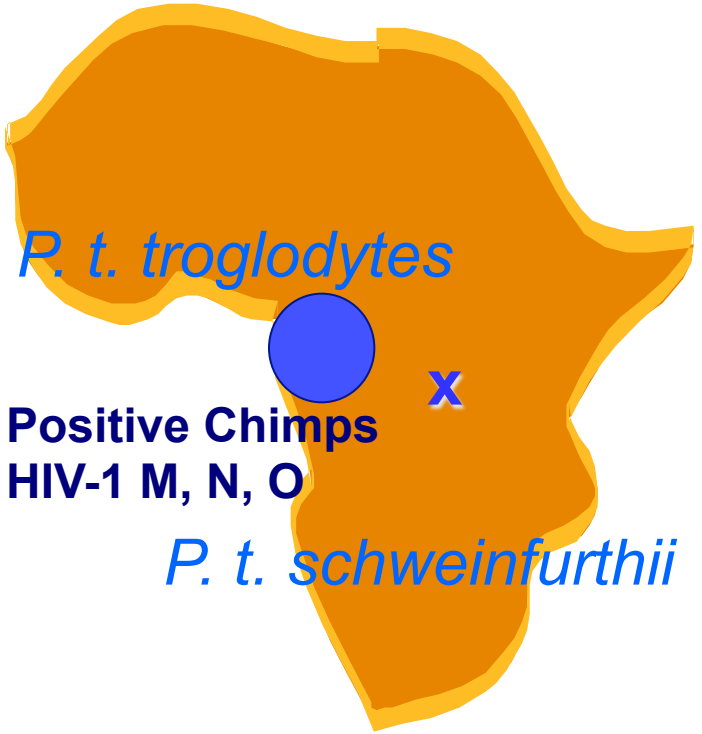
SIV/PLV Alignments

- Any non-human lentivirus is a SIV (or primate lentivirus), not just the SIV-SMM/SIV-MAC group from Sooty mangabeys.
- HIV-1s (M, N, O and P groups) are related to the SIV-CPZs from the chimps (*P. t. troglodytes)* and SIV-GORs from gorillas. We describe these alignments as HIV-1/CPZ.
- HIV-2s and SIV-MACs are related to SIV-SMMs from Sooty mangabeys. We describe these alignments as HIV-2/SMM.
- Dozens of other diverse non-human primates, such as African green monkeys, carry species-specific SIVs.
- Alignments of the diverse SIVs, plus HIVs, can help to identify highly conserved codons and other features. We describe these alignments as “other SIV” or HIV-1/HIV-2/SIV.

Primate Lentiviruses

Alignments: http://www.hiv.lanl.gov/content/hiv-db/ALIGN_CURRENT/ALIGN-INDEX.html





Van Heuverswyn, Nature 2006
Keele, *Science* 2006
Corbet, *J. Virol* 2000
Foley, HIV database

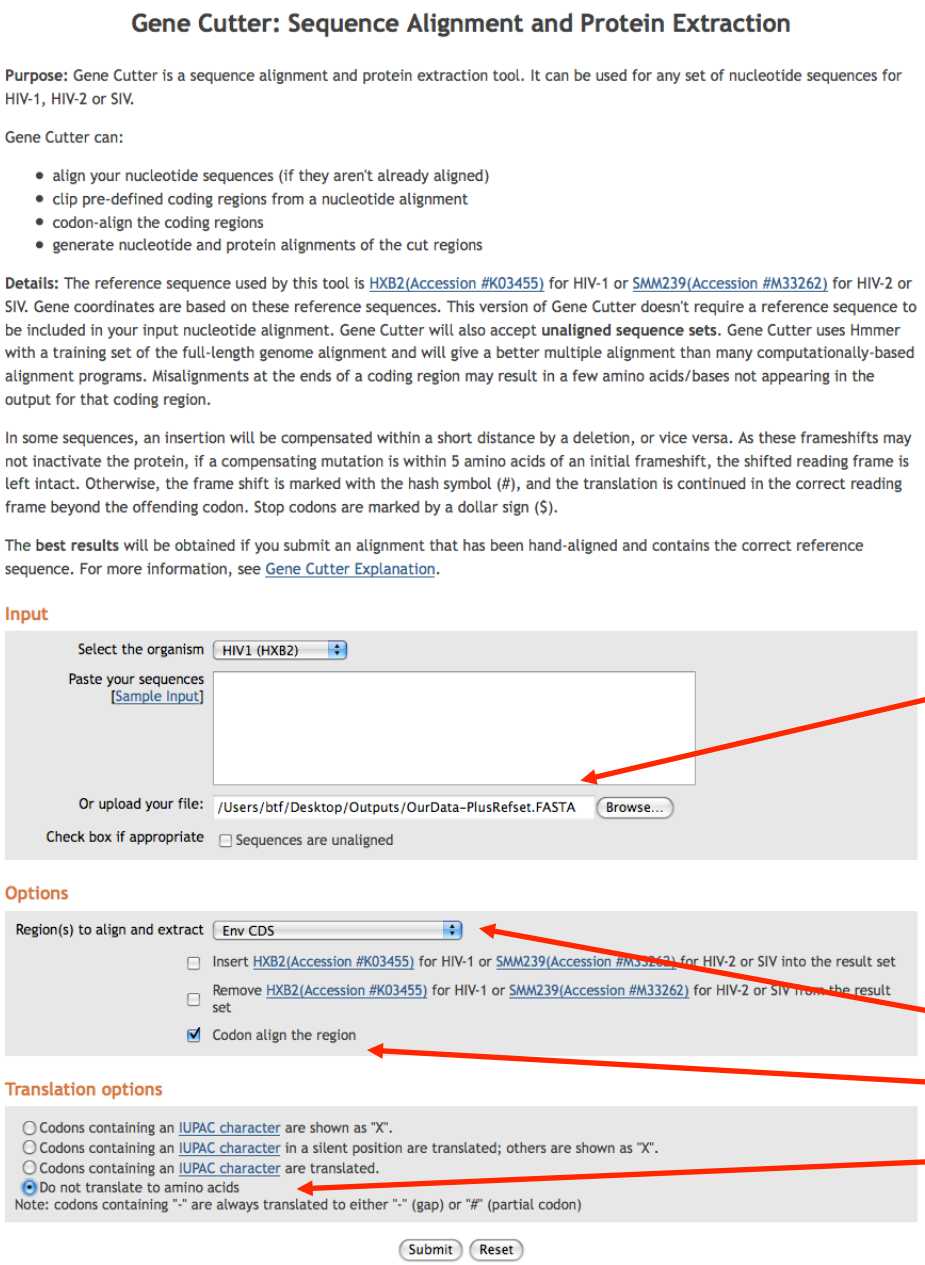
Los Alamos
NATIONAL LABORATORY

Gene Cutter

- Unconventional Alignment/Homology program
- “Cuts out” specified genes and proteins from sets of DNA sequences
  - Aligns to HXB2 via HMMer (or to SIV-Mac239 for HIV-2 and SIV-SMM)
  - Splits input sequences into genes, if desired
  - Aligns DNA sequences by codon, and translates them (including interpretation of IUPAC codes such as R for purine)
- Useful for processing new sequence data
  - annotating full length genomes
  - pulling out regions of interest from raw sequence data
- For each gene/region, maintains a list of anomalies
  - stop codons
  - codons containing multi-state characters
  - codons containing indels
- Input sequences may be aligned or unaligned
- Results may be better if the HXB2 sequence is included as a reference in your input file

Los Alamos
NATIONAL LABORATORY

GeneCutter



Gene Cutter: Sequence Alignment and Protein Extraction

Purpose: Gene Cutter is a sequence alignment and protein extraction tool. It can be used for any set of nucleotide sequences for HIV-1, HIV-2 or SIV.

Gene Cutter can:

- align your nucleotide sequences (if they aren't already aligned)
- clip pre-defined coding regions from a nucleotide alignment
- codon-align the coding regions
- generate nucleotide and protein alignments of the cut regions

Details: The reference sequence used by this tool is HXB2(Accession #K03455) for HIV-1 or SMM239(Accession #M33262) for HIV-2 or SIV. Gene coordinates are based on these reference sequences. This version of Gene Cutter doesn't require a reference sequence to be included in your input nucleotide alignment. Gene Cutter will also accept **unaligned sequence sets**. Gene Cutter uses Hmmer with a training set of the full-length genome alignment and will give a better multiple alignment than many computationally-based alignment programs. Misalignments at the ends of a coding region may result in a few amino acids/bases not appearing in the output for that coding region.

In some sequences, an insertion will be compensated within a short distance by a deletion, or vice versa. As these frameshifts may not inactivate the protein, if a compensating mutation is within 5 amino acids of an initial frameshift, the shifted reading frame is left intact. Otherwise, the frame shift is marked with the hash symbol (#), and the translation is continued in the correct reading frame beyond the offending codon. Stop codons are marked by a dollar sign ($).

The **best results** will be obtained if you submit an alignment that has been hand-aligned and contains the correct reference sequence. For more information, see Gene Cutter Explanation.

Input

Select the organism: HIV1 (HXB2)

Paste your sequences [Sample Input]

Or upload your file: /Users/btf/Desktop/Outputs/OurData-PlusRefset.FASTA Browse...

Check box if appropriate: Sequences are unaligned

Options

Region(s) to align and extract: Env CDS

Insert HXB2(Accession #K03455) for HIV-1 or SMM239(Accession #M33262) for HIV-2 or SIV into the result set

Remove HXB2(Accession #K03455) for HIV-1 or SMM239(Accession #M33262) for HIV-2 or SIV from the result set

Codon align the region

Translation options

Codons containing an IUPAC character are shown as "X".
Codons containing an IUPAC character in a silent position are translated; others are shown as "X".
Codons containing an IUPAC character are translated.
Do not translate to amino acids
Note: codons containing "-" are always translated to either "-" (gap) or "#" (partial codon)

Submit Reset

Please be patient. Your input file must download to our server, where the actual work is performed. This can take several

Input is our data plus the “reference Set” and any other sequences we chose to add from the search interface.
Input: GeneCutterInput.FASTA
Output: GeneCutterOutputAll.FASTA

For this exercise, we want the Env gene, codon aligned, but not translated to proteins.
Output: GeneCutterOutputEnv.FASTA

Los Alamos NATIONAL LABORATORY

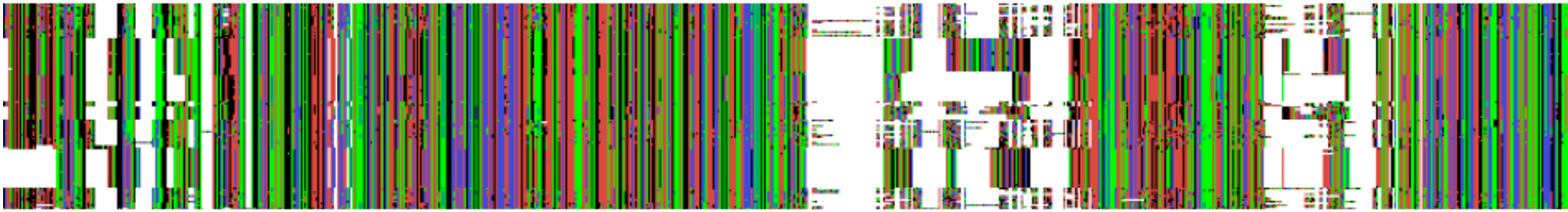
GeneCutter Results

- Results are stored on our server
  - An HTML link is e-mailed to the user when the run is complete
  - For this workshop, we will provide example.

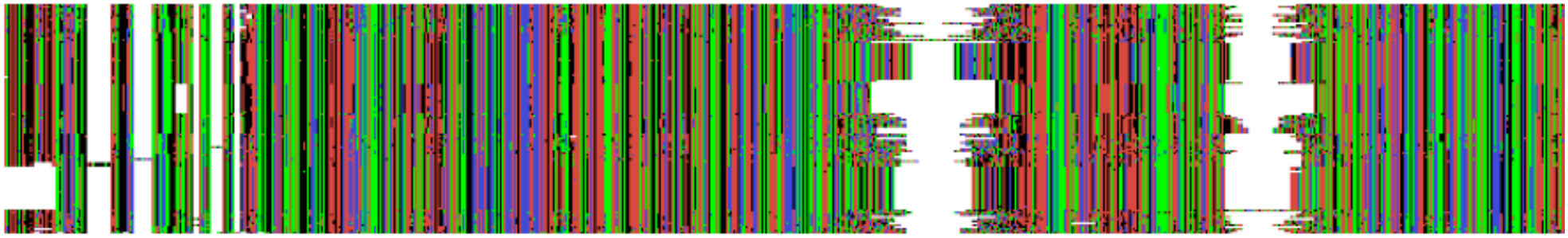
GeneCutter Result

Result saved in Outputs folder
Alignments viewed with Pixel
http://www.hiv.lanl.gov/content/sequence/pixel/pixel.html

Our data aligned to reference set by search tool:
GeneCutterInput.FASTA
(output of search and tree build was input to GeneCutter)



Our data aligned to reference set by GeneCutter:
Outputs: GeneCutterOutputENV.FASTA



Can also be viewed with BioEdit, Se-Al or other multiple sequence alignment editors.

Los Alamos
NATIONAL LABORATORY

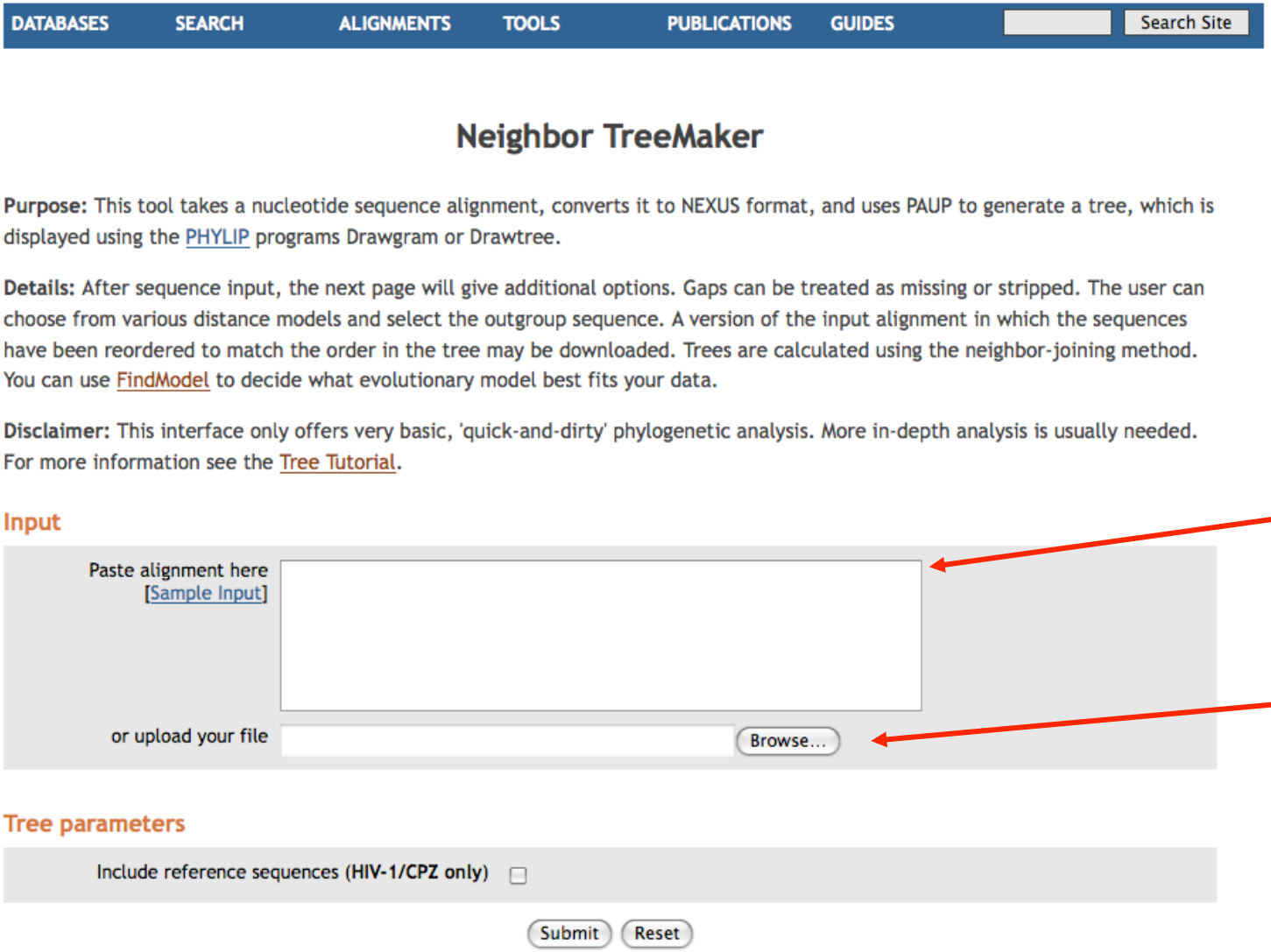
Treemaker

Check for phylogenetic relatives:

- TreeMaker produces a Neighbor Joining tree for a quick comparison
- TreeMaker uses PAUP* for its calculations; a few model options are available
- Reference sequences can be included, and are aligned to the input automatically
- Trees are displayed using PHYLIP and ATV
- The alignment used for the tree can also be downloaded
- A Phyml interface is also available

http://www.hiv.lanl.gov/content/sequence/PHYML/interface.html

http://www.hiv.lanl.gov/components/sequence/HIV/treemaker/treemaker.html

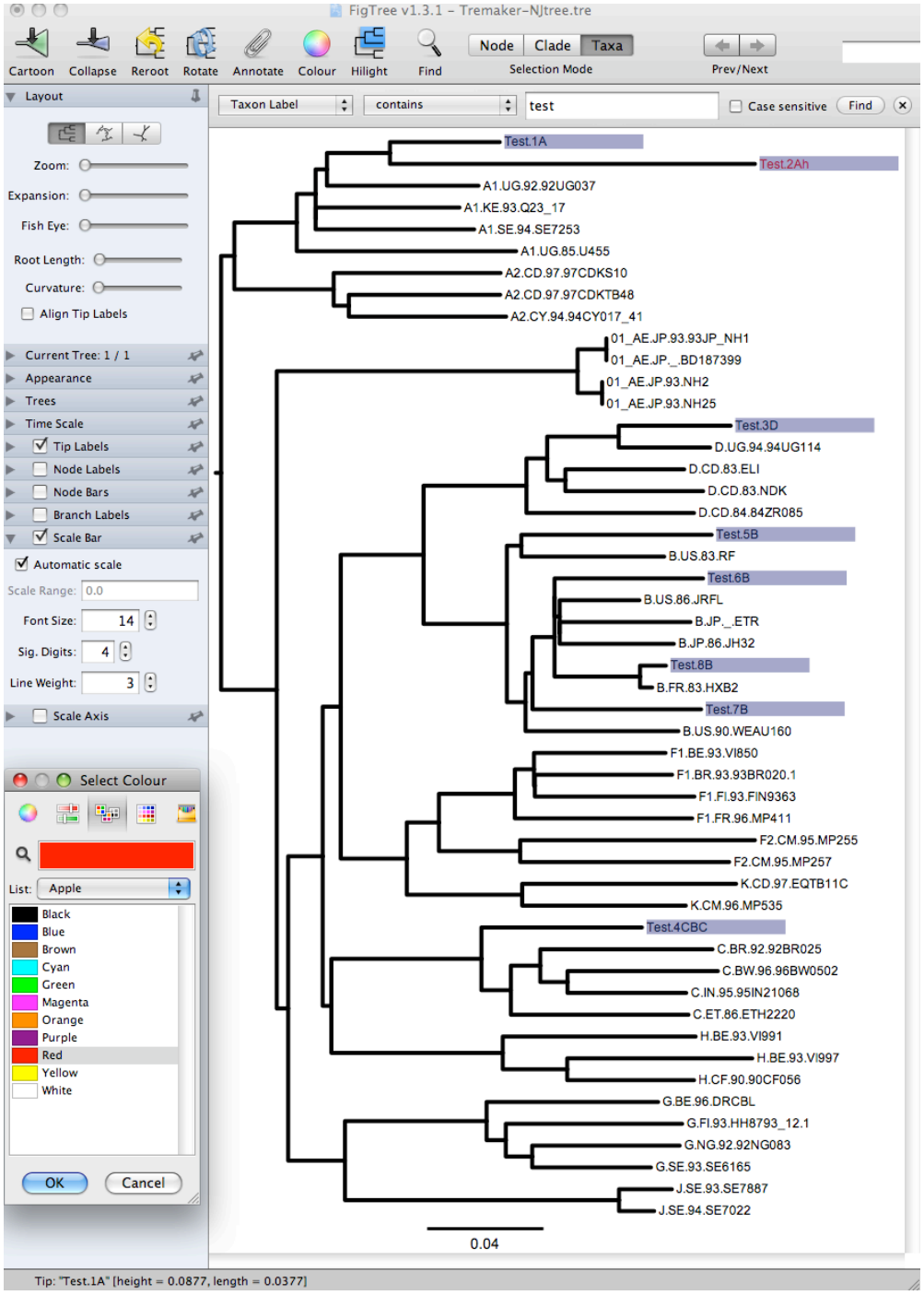


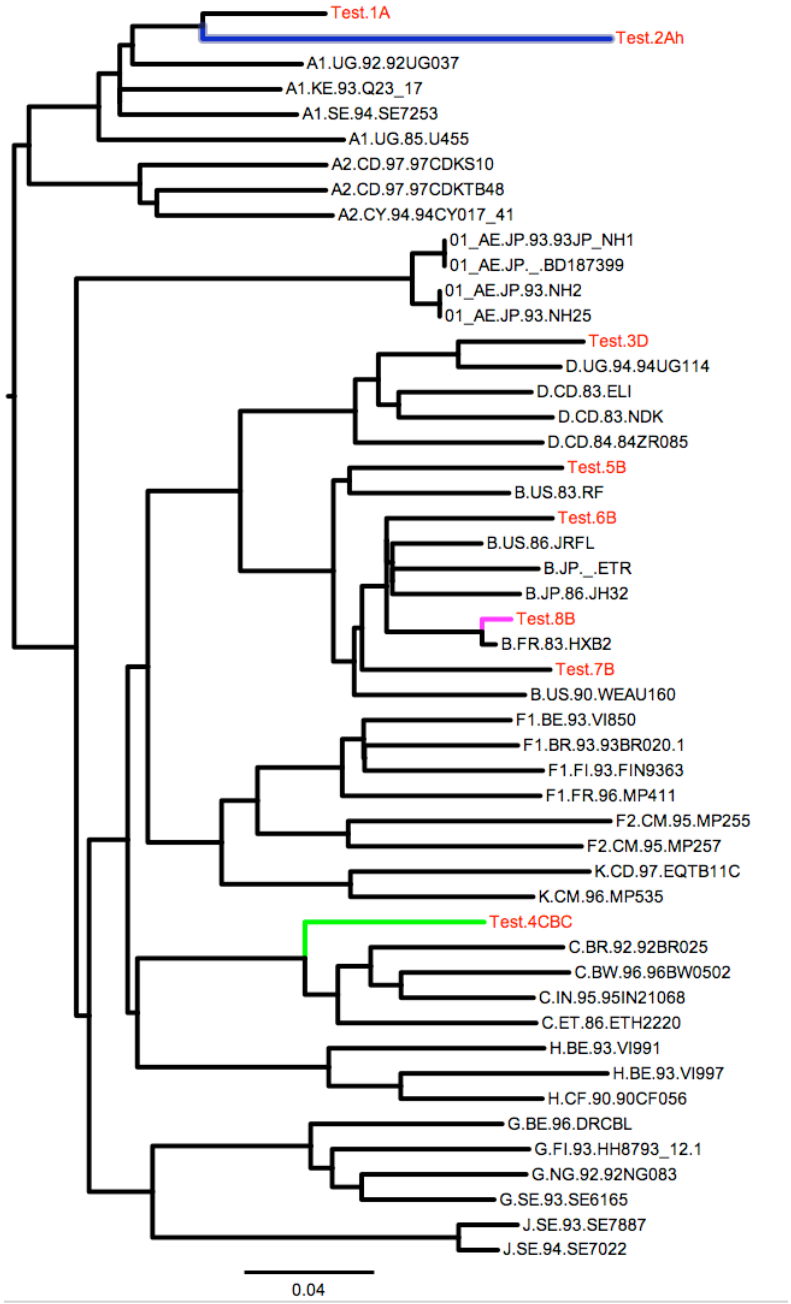
Paste or type a DNA **alignment** here.

OR upload an alignment file here.

Los Alamos NATIONAL LABORATORY

http://tree.bio.ed.ac.uk/software/figtree/





Los Alamos
NATIONAL LABORATORY

HIV/SIV Sequence Locator Tool

- Instantly computes position numbers of DNA or protein fragments relative to a reference strain (HXB2r for HIV-1, SMM239 for SIV)
  - Such numbers, often included in the literature, are frequently incorrect
- Shows the location of the sequence on an HIV map
- Presents protein translations of DNA sequences
- Can be used for input into the search interface, to align a new sequence you have generated with the database set
- Can also retrieve reference sequences
  - by coordinates (range of base or amino-acid positions)
  - by single position (retrieves flanking sequences)

Los Alamos
NATIONAL LABORATORY

http://www.hiv.lanl.gov/content/sequence/LOCATE/locate.html

HIV Sequence Locator Tool

Purpose: This tool has several purposes. It can find the start and end coordinates (relative to the reference strain HXB2) of your input sequence(s) and show which genes or proteins it covers, along with a graphical view of the location of your sequence(s) relative to the reference sequence. The tool will display both the nucleotide sequence and protein translation of your input as it aligns to HXB2. It will also check the reverse complement of your input sequence, and report the orientation with the best match. Another use is to retrieve a section of the HXB2 reference sequence based on its coordinates.

How to use: To find the coordinates for your sequence, either upload or paste your sequence (any format) in the box below, or (for database sequences only) enter GenBank accession numbers. To retrieve the HXB2 sequence for a set of coordinates (see HIV coordinate map), enter the coordinates and choose the region. To retrieve the entire gene or protein, enter coordinate values of "1" and "end". To retrieve a single nucleotide or range with its surrounding 42-nucleotide sequence, enter the single coordinate in the "from" field and check the box. For more details, see Sequence Locator Explanation.

Useful Links:
HXB2 numbering | SIVmm239 numbering (review articles)
HXB2 spreadsheet | SIVmm239 spreadsheet (spreadsheets with base-by-base annotation)

Find the location of a sequence

Sequence type: (•) Let program decide ( ) HIV ( ) SIV

Paste your input here [Sample Input]

or upload your file [Browse...]

-- OR --

Retrieve a region by its coordinates

Enter coordinates: from [ ] to [ ] (Enter '1' and 'end' to retrieve the entire region.)

Region [Complete]

Retrieve (•) Nucleotide or ( ) protein output

[ ] include surrounding region

[Submit] [Reset]

Paste or type a DNA or protein sequence here.

OR enter numeric coordiantes here.

Los Alamos NATIONAL LABORATORY

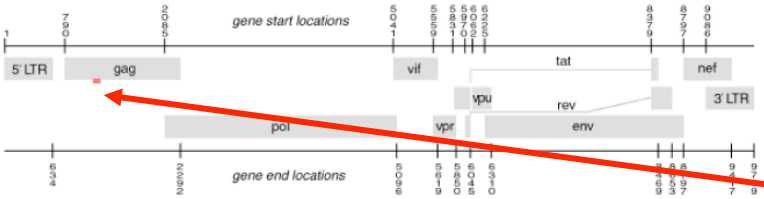
Sequence Locator: “find my sequence”

Sequence Location Result

Organism: **HIV-1**

Sequence seq1

LOCATION from start of HXB2 genome 1162 → 1251 (shown as red bar in map between reading frames 1 and 2)



Align Query

Result for Sample Input DNA Query sequence

Location in genome mapped in red.

Table of genomic regions touched by query sequence. (Protein translation of query shown in blue.)				
CDS	Nucleotide position relative to CDS start in HXB2	Nucleotide position relative to query sequence start	Nucleotide position relative to HXB2 genome start	**Amino Acid** position relative to protein start in HXB2
Gag	373 → 462	1 → 93	1162 → 1251	125 → 154
Notice: length of Gag portion of query is greater than its length in HXB2.				
SNQMVSQNCPIVQNIQGQVVHQAISPRTLNA				
p17	373 → 396	1 → 27	1162 → 1185	125 → 132
Notice: length of p17 portion of query is greater than its length in HXB2.				
SNQMVSQNC				
p24	1 → 66	28 → 93	1186 → 1251	1 → 22
PIVQNIQGQVVHQAISPRTLNA				

Alignment of the query sequence to HXB2 (Similarity 94.6%):

```
Query AGCAATCAGA TGGTCAGCCA AAATTGCCCT ATAGTGCAGA ACATCCAGGG  50
      :::::::::    :::::::::: ::::: :::: :::::::::: ::::::::::
 HXB2 AGCAATCA-- -GGTCAGCCA AAATTACCCT ATAGTGCAGA ACATCCAGGG  1208

Query GCAAGTGGTA CATCAGGCCA TATCACCTAG AACTTTAAAT GCA  93
      :::: ::::: :::::::::: :::::::::: :::::::::: :::
 HXB2 GCAAATGGTA CATCAGGCCA TATCACCTAG AACTTTAAAT GCA  1251
```

Numeric coordinates useful for entry on search form

DNA and protein sequence displayed

Retrieve a region by its coordinates

Enter coordinates: from [ ] to [ ] (Enter '1' and 'end' to retrieve the entire region.)

Region [Complete]

Retrieve (•) Nucleotide or ( ) protein output

[ ] include surrounding region

Submit Reset

Los Alamos NATIONAL LABORATORY

Sequence Locator:
"Retrieve from coordinates"

Table of genomic regions touched by query sequence. Query protein translation in blue.				
CDS	**NA** position relative to CDS start in HXB2	**NA** position relative to query sequence start	**NA** position relative to HXB2 genome start	**AA** position relative to protein start in HXB2
Gag	352 -> 483	1 -> 132	1141 -> 1272	118 -> 161
AAADTGHSNQVSQNYPIVQNIQGQMVHQAISPRTLNAWVKVVEE				
p17	352 -> 396	1 -> 45	1141 -> 1185	118 -> 132
AAADTGHSNQVSQNY				
p24	1 -> 87	46 -> 132	1186 -> 1272	1 -> 29
PIVQNIQGQMVHQAISPRTLNAWVKVVEE				

Sequence below includes up to 42 bases of context surrounding query sequence.

Reference Strain	*Type*	*Region*	*Start*	*End*
HXB2	nuc	complete	1141	1272

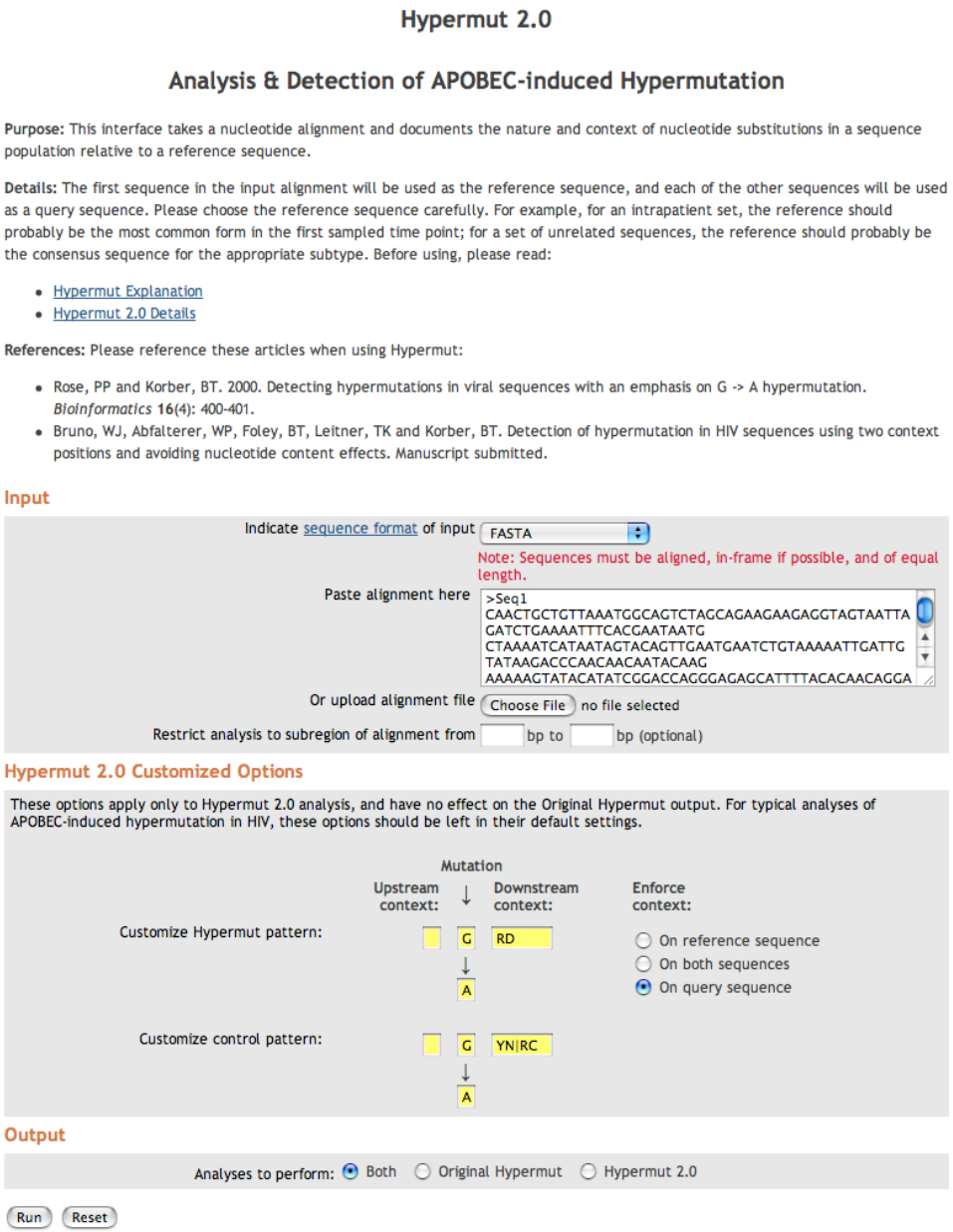
Retrieved Sequence:

GCAGCAGCTGACACAGGACACAGCAATCAGGTCAGCCAAAATTACCCTATAGTGCAGAACATCCAGGGGCAAATGGTACA
TCAGGCCATATCACCTAGAACTTTAAATGCATGGGTAAAAGTAGTAGAAGAG

Organism: HIV

Los Alamos NATIONAL LABORATORY

Hypermutation



Hypermut 2.0

Analysis & Detection of APOBEC-induced Hypermutation

Purpose: This interface takes a nucleotide alignment and documents the nature and context of nucleotide substitutions in a sequence population relative to a reference sequence.

Details: The first sequence in the input alignment will be used as the reference sequence, and each of the other sequences will be used as a query sequence. Please choose the reference sequence carefully. For example, for an intrapatient set, the reference should probably be the most common form in the first sampled time point; for a set of unrelated sequences, the reference should probably be the consensus sequence for the appropriate subtype. Before using, please read:

- Hypermut Explanation
- Hypermut 2.0 Details

References: Please reference these articles when using Hypermut:

- Rose, PP and Korber, BT. 2000. Detecting hypermutations in viral sequences with an emphasis on G -> A hypermutation. *Bioinformatics* **16**(4): 400-401.
- Bruno, WJ, Abfalterer, WP, Foley, BT, Leitner, TK and Korber, BT. Detection of hypermutation in HIV sequences using two context positions and avoiding nucleotide content effects. Manuscript submitted.

Input

Indicate sequence format of input: FASTA

Note: Sequences must be aligned, in-frame if possible, and of equal length.

Paste alignment here:
>Seq1
CAACTGCTGTTAAATGGCAGTCTAGCAGAAGAAGAGGTAGTAATTA
GATCTGAAAATTTCACGAATAATG
CTAAAATCATAATAGTACAGTTGAATGAATCTGTAAAAATTGATTG
TATAAGACCCAACAACAATACAAG
AAAAAGTATACATATCGGACCAGGGAGAGCATTTTACACAACAGGA

Or upload alignment file: Choose File — no file selected

Restrict analysis to subregion of alignment from ____ bp to ____ bp (optional)

Hypermut 2.0 Customized Options

These options apply only to Hypermut 2.0 analysis, and have no effect on the Original Hypermut output. For typical analyses of APOBEC-induced hypermutation in HIV, these options should be left in their default settings.

Mutation — Upstream context: ↓ Downstream context: — Enforce context:

Customize Hypermut pattern: [ ] G → A, [RD] — On reference sequence / On both sequences / On query sequence

Customize control pattern: [ ] G → A, [YN|RC]

Output

Analyses to perform: Both / Original Hypermut / Hypermut 2.0

Run Reset

- Detects APOBEC related A->G hypermutation as default
- Can be adapted to detect any fuzzy motif in relation to a control pattern

Los Alamos NATIONAL LABORATORY

Hypermut results

Hypermut 2.0

Your pattern definitions are as follows. Where there is no pattern (i.e., just '...') all sequences will match.

Pattern	Upstream	From →To	Downstream
'Mut'	...	G →A	RD ...
'Control'	...	G →A	YN|RC ...

Results

'Potential Mut' or 'Potential Control' means a match to the corresponding Upstream, From, and Downstream patterns above, while an actual 'Mut' matches those and the To pattern as well. We consider a P-value less than 0.05 to indicate a hypermutant when using the default patterns.

Sequence: (Select for graphing)	Muts: (Match Sites)	Out of: (Potential Mut Sites)	Controls: (Control Muts)	Out of: (Potential Controls)	Rate Ratio: (A/B)/(C/D)	Fisher Exact P-value: (=P(Muts,Poten.Muts-Muts, Cntrls,Poten.Cntrls-Cntrls))
☑ Seq2	0	71	0	54	undef	1
☑ Seq5	4	69	1	52	3.01	0.282669
☑ Seq7	26	71	1	54	19.77	5.35061e-07
☑ Seq14	48	71	9	54	4.06	8.26961e-09

High ratio of G -> A vs. A -> G indicates hypermutation

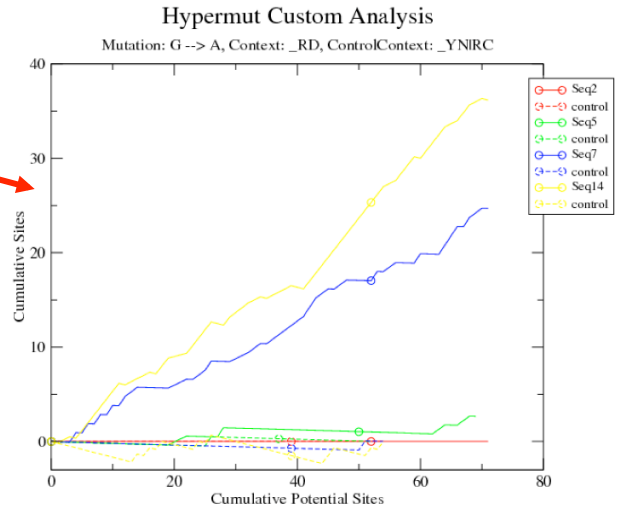
View Sites Along Sequence

Type of graph:
- ◉ Locations of Matches
- ○ Cumulative Matches (try me!)

Graph Matches (opens in a new window)

Optional Controls:
Show region: From [ ] to [ ]
Graph Title: [Hypermut Custom Analysis]
Access xmgrace compatible datafile.

Cumulative mutation Graph is useful



Original Hypermut Output

The input file has 5 sequence(s) Download the following info as text file
Sequence Length: 645
Compared to SEQ1, 264 As, 126 Gs, 92 Cs, 163 Ts

GRAPH	TABLE	Sequence names	Ratio	#diffs	perc_Gs	#A->G	#G->A	Dinuc Context: GG	GA	GC	GT	OBSERVED CHANGES
○	○	SEQ2	0/0	1	0.00	0	0	0	0	0	0	TC
○	○	SEQ5	5/2	33	3.97	2	5	2	3	0	0	GA TA GA CT GA GA AT CT AG A- T- A- G- T-

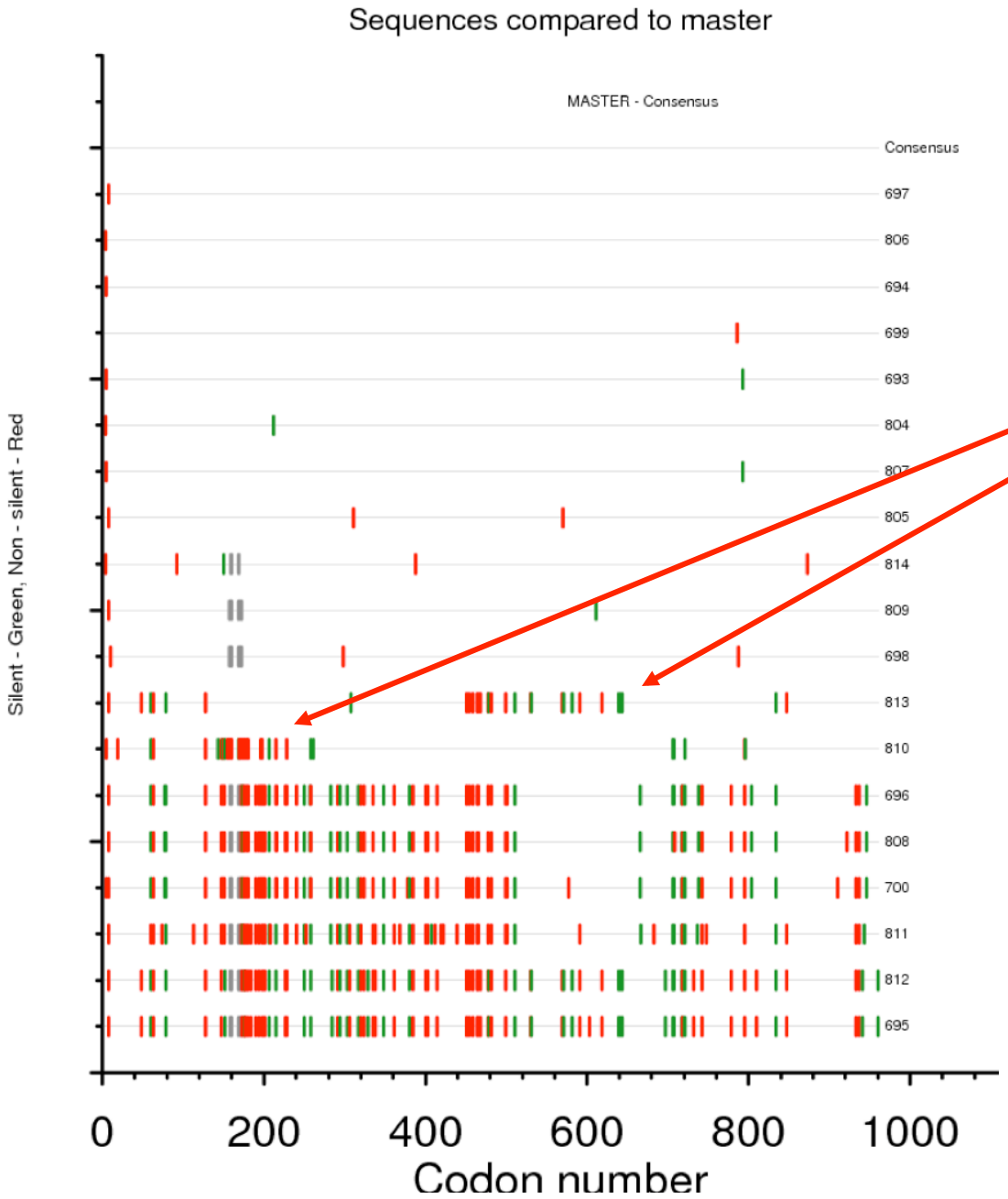
utput:
df version here.
ostscript version here.

Los Alamos NATIONAL LABORATORY

Highlighter

- Highlights mutations relative to a reference strain, particularly useful for intra-patient analyses.
- Highlights:
  - syn/non-syn
  - transition/transversion
  - Apobec motifs
- Sorts on similarity
- Visualize recombination of closely related sequences

Silent and Non-silent statistics



Nonrandom distribution of mutations evident.

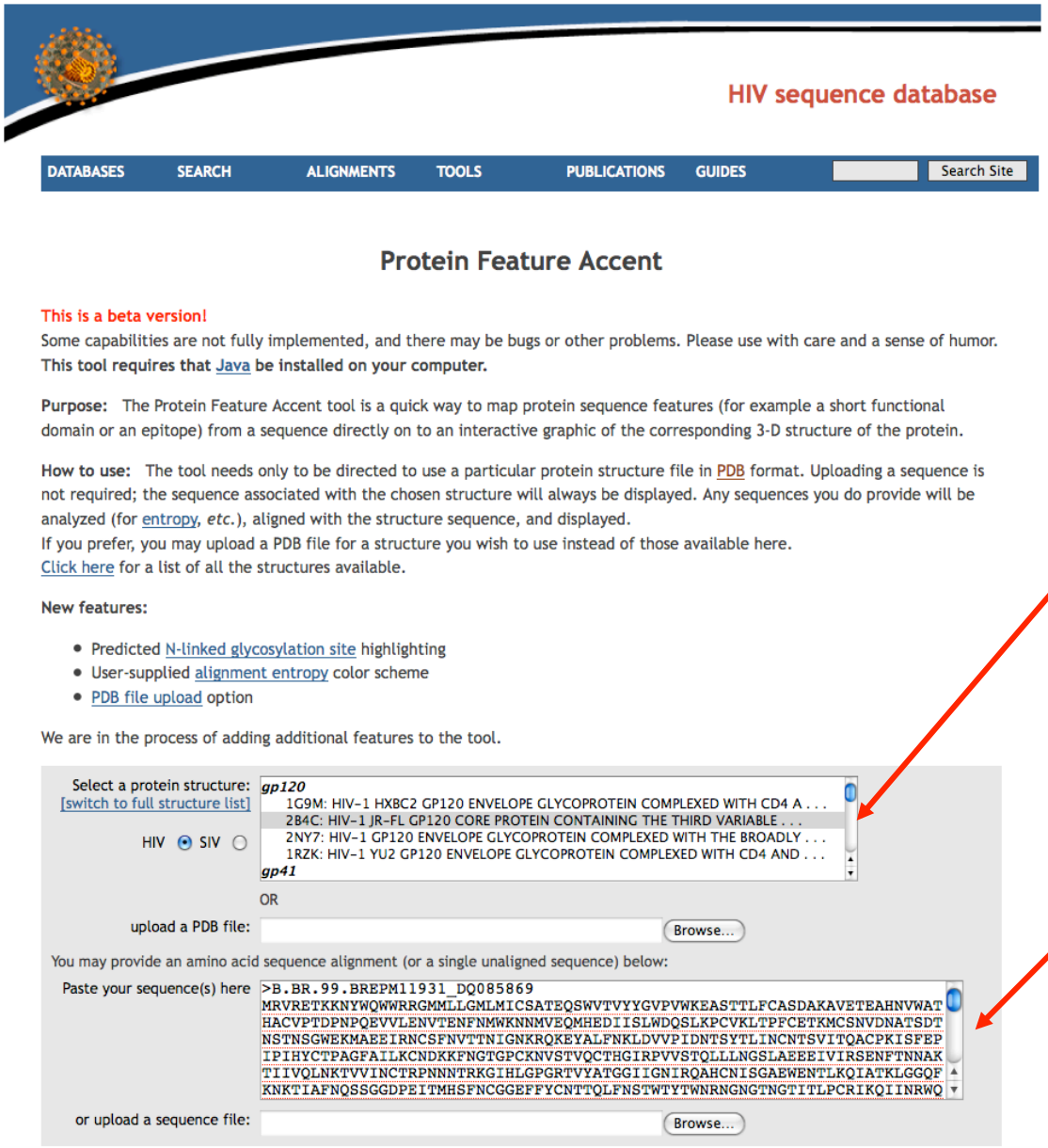
Sample Set is from a possible dual Infection, with intra-subtype recombinants evident.

Los Alamos NATIONAL LABORATORY

Protein Feature Accent

- Highlights region of interest in an HIV structure
- You can upload a PDB structure, or use one of our annotated Env structures
- You can upload your own alignment and get an entropy map

http://www.hiv.lanl.gov/content/sequence/PROTVIS/html/protvis.html

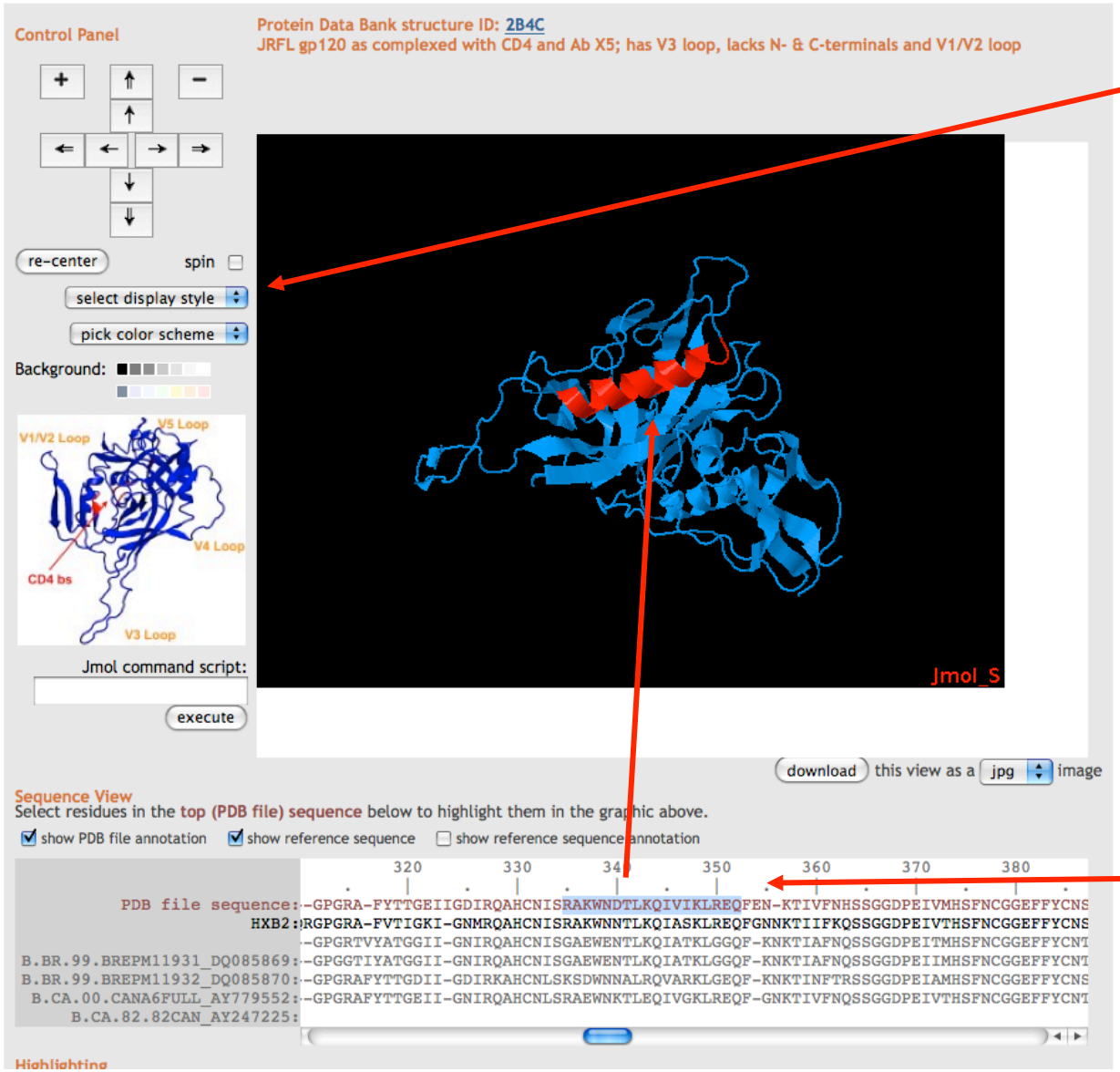


List of “recommended” PDB entries

Only a gp120 alignment is provided so far. We hope to add others. You can paste in your own.

Los Alamos NATIONAL LABORATORY

http://www.hiv.lanl.gov/content/sequence/PROTVIS/html/protvis.html



Many display options in JMol are "built in" to this web tool.
Use the JMol command script box below for other commands.

One of the color schemes is "color by entropy" based on divrsity in the alingment added below.

Selected region gets highlighted in structure

Los Alamos NATIONAL LABORATORY

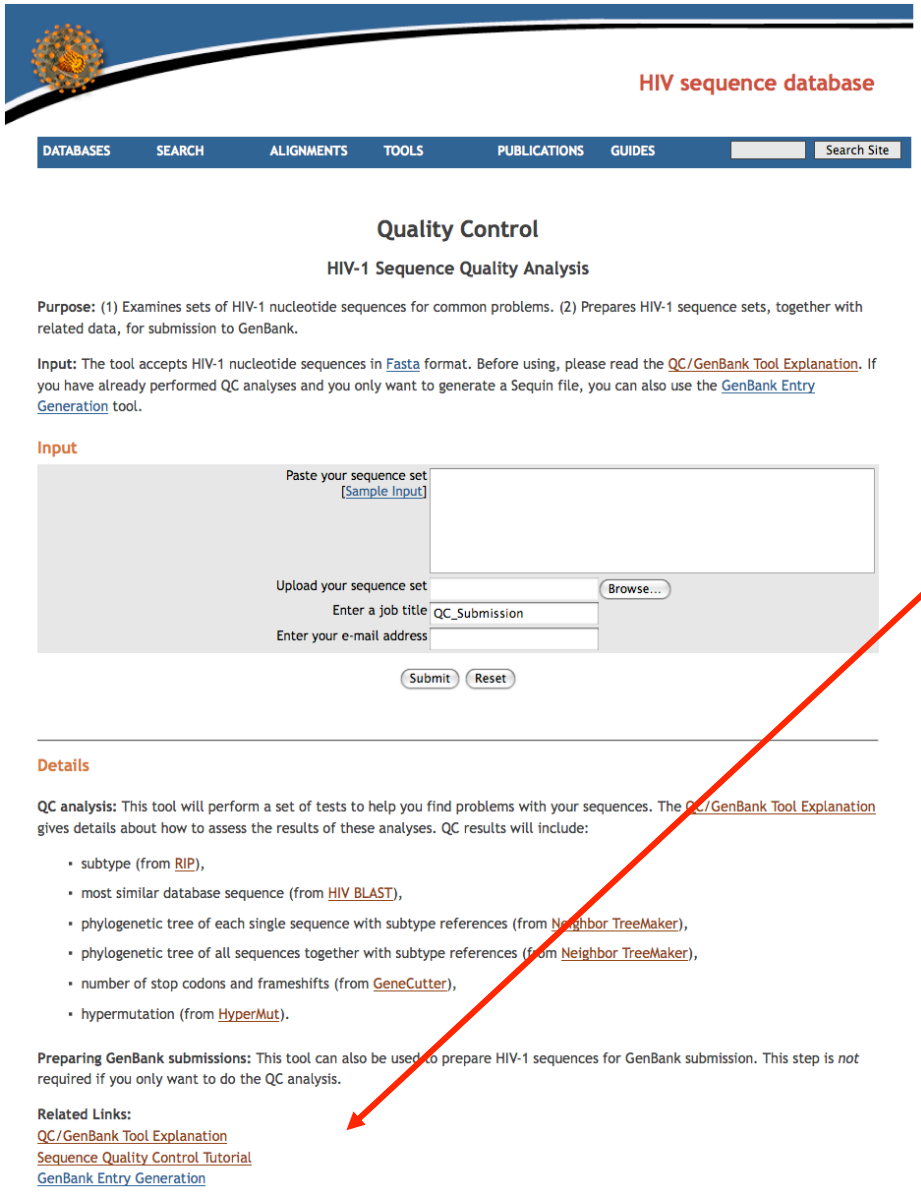
Quality Control Tool

- Built from existing HIV database tools
- GeneCutter
- RIP
- Hypermut
- Neighbor-joining Trees
- Output is an email containing a link to a summary report
- http://www.hiv.lanl.gov/content/hiv-db/QC/index.html (beta version)

Los Alamos
NATIONAL LABORATORY

Quality Control Tool

http://www.hiv.lanl.gov/content/sequence/QC/index



Recently added shortcuts to GenBank entry creation tool.

Requires FASTA format sequences, and a comma separated values (CSV) file of annotations, as described on the help page.

http://www.hiv.lanl.gov/content/sequence/QC/field_help.html

Easy to enter in spreadsheet like EXCEL, and then export as CSV format.

Los Alamos NATIONAL LABORATORY

Quality Control Tool

- Summary of results from analysis programs
- Click on each result to obtain full analysis
- Useful for helping to determine subtype, hypermutation, mislabeling of samples

HIV Database Workshop

www.hiv.lanl.gov
seq-info@lanl.gov

Presenters: Brian Foley, Karina Yusim

Database PIs: Bette Korber, Thomas Leitner, Karina Yusim

Additional database staff: Werner Abfalterer, Will Fischer, Peter Hraber, Elisabeth Sharon Fung, Robert Funkhouser, Jenni Macke, Kumkum Ganguly, James Szinger, and Hyejin Yoon

Project Officer: Stuart Shapiro, NIAID, NIH

Editors: Christian Apetrei, Beatrice Hahn, Ilene Mizrachi, James Mullins, Andrew Rambaut, Steve Wolinsky, Dan Barouch, Christian Brander, Rob De Boer, Bart Haynes, Richard Koup, John Moore, Bruce Walker, David Watkins

Theoretical Biology and Biophysics, T-6
Los Alamos National Laboratory

Presenters:

Brian Foley: Responsible for HIV Sequence Database and Vaccine Trials Database content and research; has a background in bioinformatics, lentivirus evolution and virology.

Karina Yusim: Co-PI of the project; Primary Editor for the HIV Immunology Database, responsible for HIV Immunology Database website and tools; has a background in sequence analysis, bioinformatics, immunology and applied mathematics.

seq-info@lanl.gov

Los Alamos NATIONAL LABORATORY

Workshop Topics

HIV Sequence Database and Immunology Database
Brian Foley, Karina Yusim

Session 2

Tuesday,
Feb 12
11:00 – 12:30

Immunology database introduction
Epitope maps and epitope summary tables
T-cell epitope search
T-cell epitope variants
Antibody search
List of most broadly neutralizing antibodies
HIV/SIV sequence locator tool
QuickAlign – Align an epitope to the database alignments
Motifscan – find HLA anchor residues in a protein
N-glycosite – finds N-linked glycosylation sites
ELF – epitope location finder
Peptgen – list peptides for reagent development

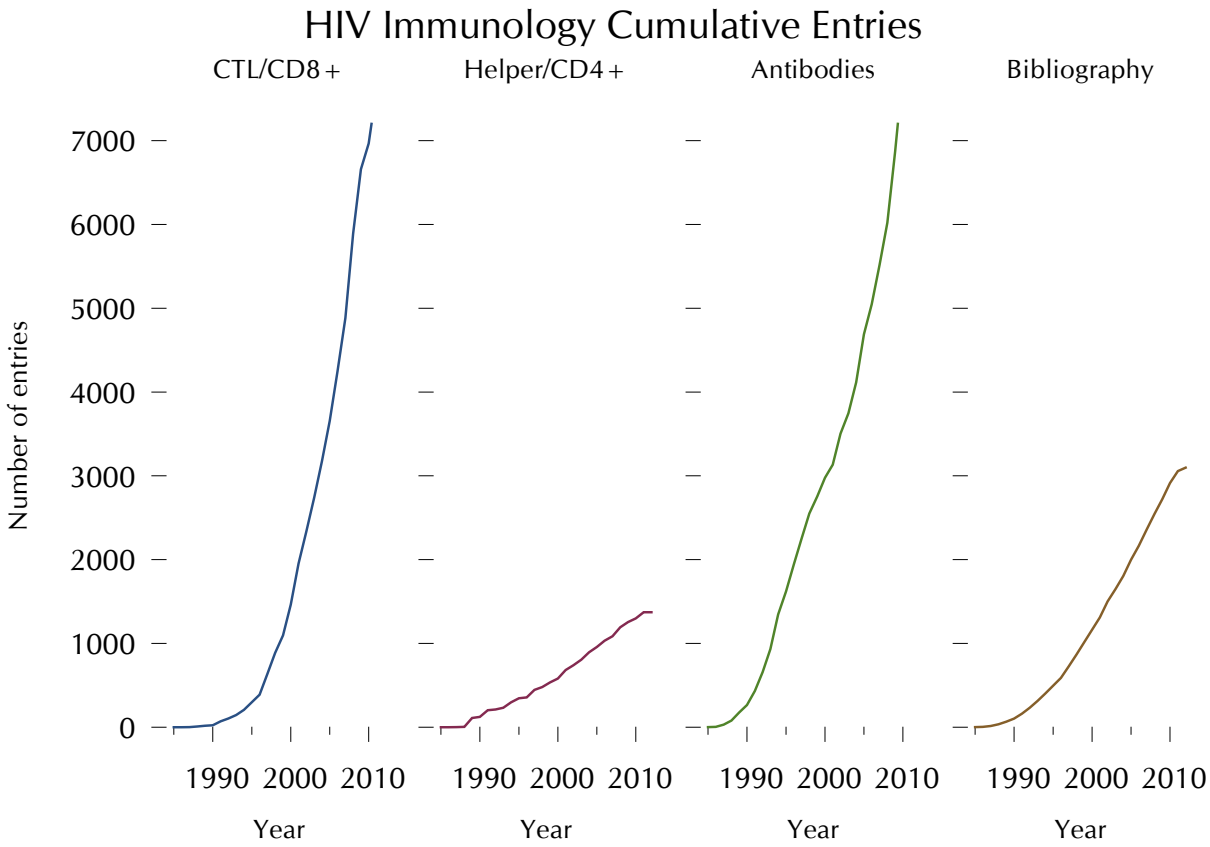
Mosaic Vaccine Maker, Epicover, and Posicover
- generate candidate vaccines
- estimate epitope coverage
- determine regional epitope coverage

Immunology Database Overview

- Incorporates published HIV T cell (CTL, T-helper) epitope and Antibody information (emphasis on monoclonals)
- Key information regarding what is learned about epitopes and mAbs in each paper is included
- Types of data recorded:
  - Epitope sequence and location: HXB2 numbering, subtype
  - Natural infection or vaccine
  - Host HLA or MHC
  - Ab isotype, binding region, species
  - Notes summarize main findings

Immunology Database Statistics

- Contents: data from 1985 through 2013
  - 7441 CTL entries
  - 1315 T-helper entries
  - 2386 Ab entries
  - 3090 published citations



- Usage:
  - ~ 70% papers entered in CTL epitope database use HIV Immunology Database resources

Los Alamos NATIONAL LABORATORY

HIV Molecular Immunology Database

The HIV Molecular Immunology Database is an annotated, searchable collection of HIV-1 cytotoxic and helper T-cell epitopes and antibody binding sites.

Search the Molecular Immunology Database

- CTL/CD8+ Search
- T Helper/CD4+ Search
- Antibody Search
- Search Help

Database Products

- All Database products and publications
- Epitope maps
- Epitope summary tables
- Epitope alignments
- Epitope variants and escape mutations
- The HIV Molecular Immunology Compendium
- About the HIV Molecular Immunology Database
- How to cite this database

Tools and Data Sets

- Tools & Links for immunologists
- HIV "A list" CTL/CD8+ Epitopes (PDF) review article summarizing the best-characterized HIV epitopes
- SIV Epitopes (PDF) review article summarizing known SIV epitopes
- Identifying HLA-Associated Polymorphisms in HIV-1 (PDF) review article summarizing HIV polymorphism associated with escape mutations. Also a table of polymorphisms.
- HLATEM HLA Typing and Epitope Mapping Data Sets
- Standardized Assessments of Neutralizing Antibodies for HIV/AIDS Vaccine Development Assay protocols from Duke Central Reference Laboratory

Los Alamos
NATIONAL LABORATORY

Immunology Database Products

- Epitope maps (species/HLA for T cell epitopes; species/MAb name for Ab)

- Epitope summary tables:
  - All CTL and Helper epitopes and Ab binding sites
  - Variants of CTL epitopes
  - Christian Brander keeps an “A list” of HIV CD8+ T-cell epitopes – experimentally validated optimal epitopes with known HLA presenting molecules, will be updated soon
  - “B list” – a comprehensive list of all unique epitopes in the database (unknown HLA, boundaries not fully defined…)
  - All antibodies organized by protein and binding region
  - Antibody “A-list” – a table of the most broadly neutralizing MAbs, with links to sequence and structure

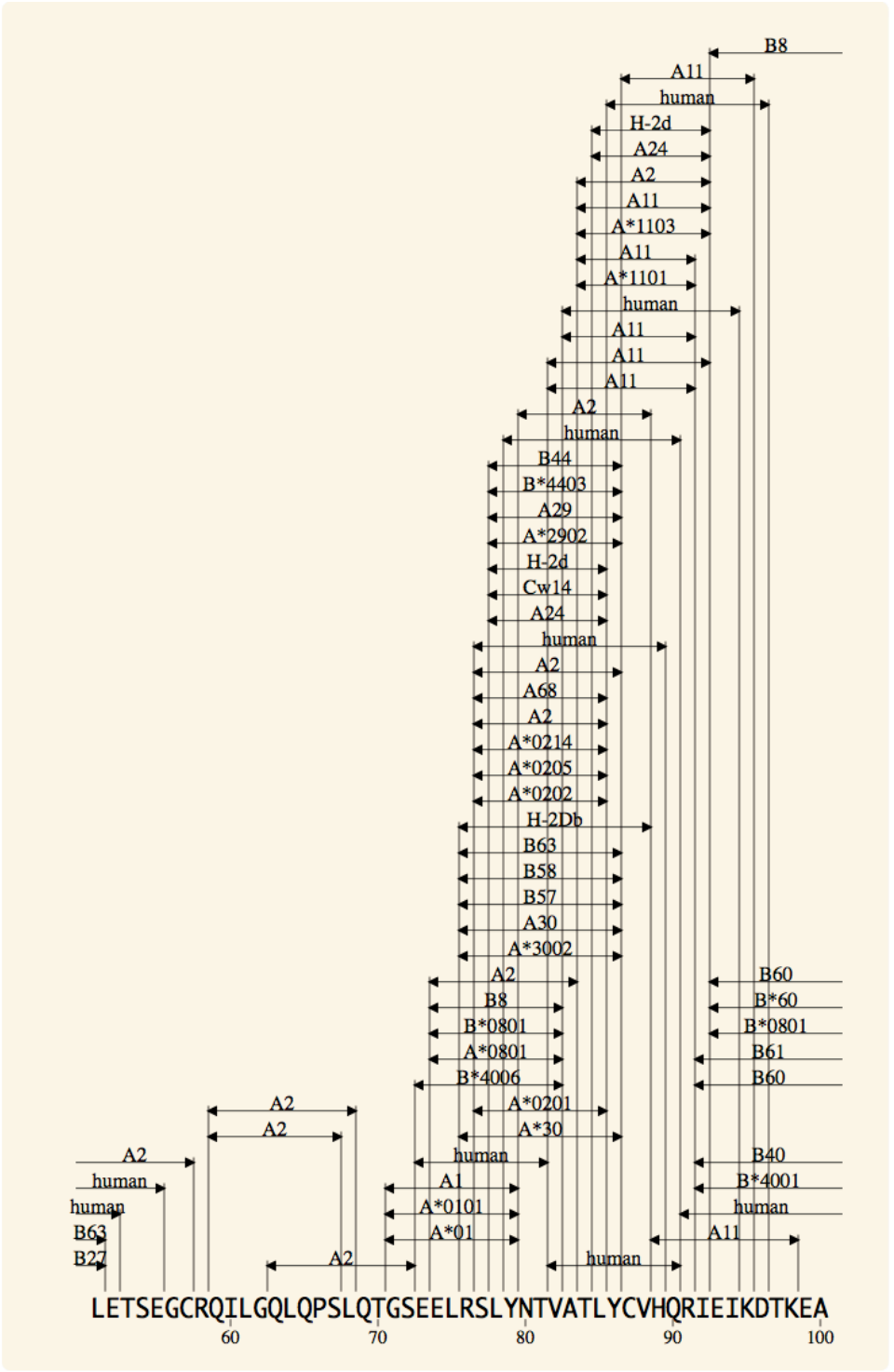
- Tools for immunologists

- Yearly HIV Molecular Immunology Compendium

Los Alamos
NATIONAL LABORATORY

p17 CTL/CD8+ Epitope Map

- Epitopes up to 14 aa long are mapped on HXB2
- HXB2 sequence may differ
- Epitopes with identical boundaries and HLA fields are included in the maps only once
- The epitope maps are interactive!



Los Alamos NATIONAL LABORATORY

CTL/CD8+ Epitope Summary (B-list)

- List of all epitopes up to 21 aa long
- Unlike epitope maps that show epitope locations, here each epitope sequence is shown

Epitope	Protein	HXB2 Location	Subtype	Species	HLA
MGARASVLSG	p17	1-10	CRF01_AE	human	
ASVLSGGEL	p17	5-13	B	human	
ASILRGGKLDK	p17	5-15	C	human	
SVLSGGQLDR	p17	6-15	B	human	A11
LSGGELDRWEK	p17	8-18		macaque	
GELDRWEKI	p17	11-19	B	human	B*4002, B40
GQLDRWEKI	p17	11-19	B	human	
GKLDSWEKIRLR	p17	11-22	A, CRF01_AE, CRF02_AG	human	
GKLDAWEKIRLR	p17	11-22	CRF01_AE	human	
ELDRWEKIRL	p17	12-21	B, C	human	B63
EKIRLRPGGKKYKL	p17	17-31		human	B27, B7
KIRLRPGGK	p17	18-26	A, A1, B, CRF01_AE	human, transgenic mouse	A*0301, A11, A3, B27, B7
KIRLRPGGKK	p17	18-27	B, C, multiple	human	A*0301, A11, A3, B27
KIRLRPGGKKKYKL	p17	18-31		human	A3, B62

Best-defined CTL/CD8+ Epitope Summary (A-list)

- Selective list of best defined epitopes as described by Christian Brander and colleagues

Epitope	Protein	HXB2 Location	Subtype	Species	HLA
GELDRWEKI	p17	11-19		human	B*4002
KIRLRPGGK	p17	18-26		human	A*0301
IRLRPGGKK	p17	19-27	B	human	B*2705
RLRPGGKKK	p17	20-28		human	A*0301
RLRPGGKKKY	p17	20-29	B	human	A*0301
GGKKKYKLK	p17	24-32	B	human	B*0801
KYKLKHIVW	p17	28-36	B	human	A*2402
HLVWASREL	p17	33-41		human	Cw*0804
LVWASRELERF	p17	34-44		human	A30
WASRELERF	p17	36-44	B	human	B*3501
ELRSLYNTV	p17	74-82		human	B*0801
RSLYNTVATLY	p17	76-86	B	human	A*3002, B58, B63
SLYNTVATL	p17	77-85	B	human	A*0201, A*0202, A*0205
LYNTVATL	p17	78-85		human	Cw14
LYNTVATLY	p17	78-86		human	A*2902, B*4403
TLYCVHQK	p17	84-91		human	A*1101
IEIKDTKEAL	p17	92-101		human	B*4001
NSSKVSQNY	p17	124-132	B	human	B*3501

Los Alamos NATIONAL LABORATORY

Immunology Database: Search

- T Cells
  - Cytotoxic T Lymphocytes (CTL)
  - Helper T Lymphocytes (T-helper)
  - Organization is identical for CTL and T-helper
  - One reference per entry, epitope/HLA combinations are often repeated
- B Cells (Antibodies)
  - One entry for each monoclonal antibodies
  - Many references per entry (> 400 for some well studied MAbs)

CTL/CD8+ T-cell Search

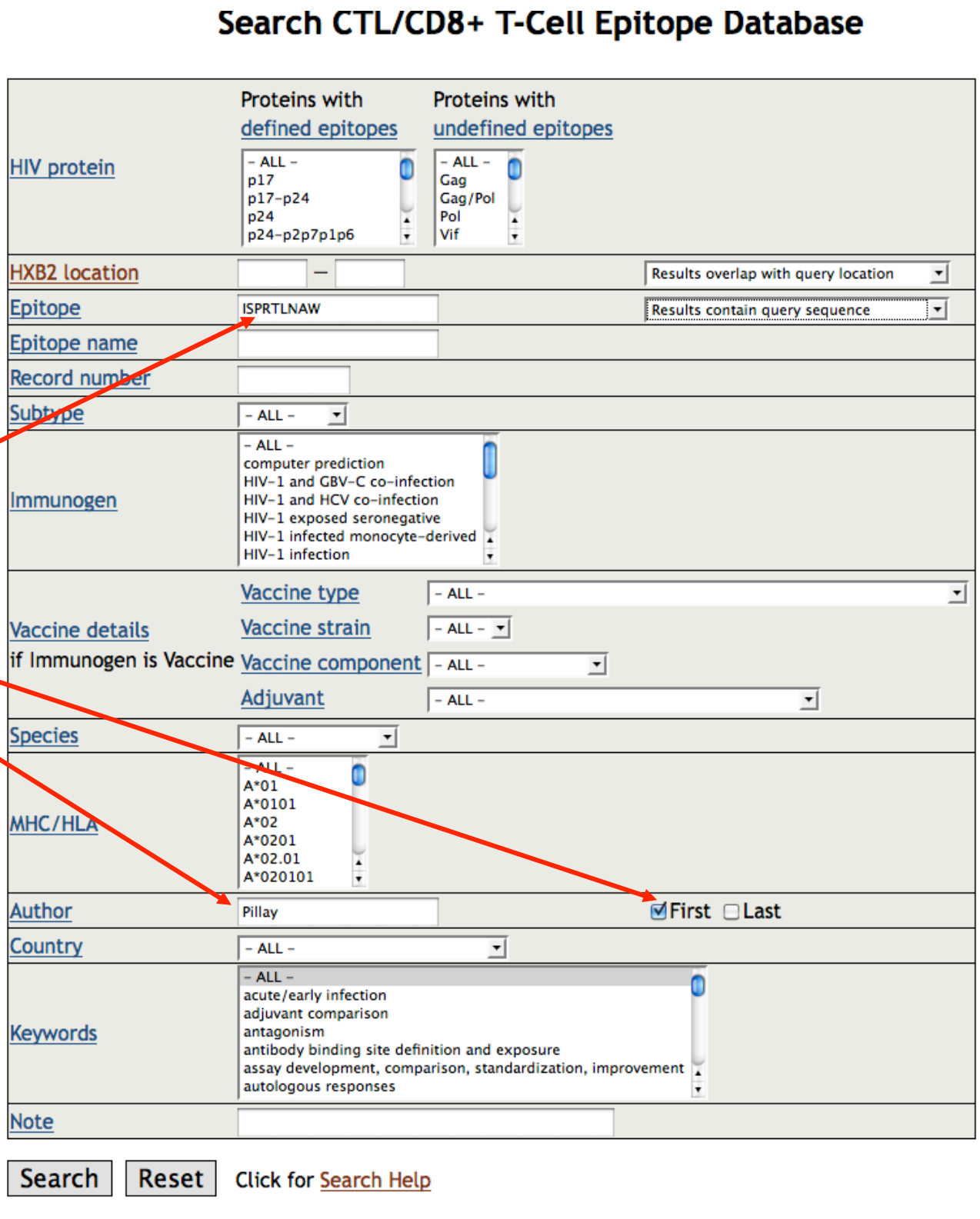
- Can search by HIV protein, Epitope Sequence, Subtype, Immunogen, Vaccine Details, Species, presenting MHC/HLA, Author, Country, Keywords
- Can now search on epitope location and find fuzzy matches, overlaps and embedded epitopes
- Search example:
  - SLYNTVATL – 254 entries
  - To narrow the search use keyword “escape” – 32 entries
- Additional information provided in the entry:
  - Location, Donor MHC/HLA, experimental methods, Notes
  - CTL epitope variants if studied in the paper
  - Link to all entries for a reference
  - PubMed links to papers
  - Link to Epitope Maps
  - Link to Epitope Alignment (Extracted from HIV-sequence database, includes subtype, country and year of sampling)

CTL/CD8+ T-cell Search

Search for

Epitope: ISPRTLNAW

First Author: Pillay



Search CTL/CD8+ T-Cell Epitope Database

Found 1 matching record:

Displaying record number 53832

HXB2 Location	p24(15-23)
Author Location	Gag(147-155)
Epitope	ISPRTLNAW
Subtype	C
Species (MHC/HLA)	human(B57)
Immunogen	HIV-1 infection
Donor MHC/HLA	A*3001, A*66, B*4201, B*5802, Cw*0602, Cw*1701; A*66, A*68, B*57, B*5802, Cw*0602, Cw*0701
Country	South Africa
Experimental methods	CD8 T-cell Elispot - IFNy
Keywords	epitope processing, responses in children, mother-to-infant transmission, escape, acute/early infection

p24 Epitope Map

Epitope Alignment

Show epitope variants

Variant details with annotator's notes

Notes

- HIV-specific CTLs in infants were shown to be able to select for viral escape variants early in life, despite a lack of escape with the same CTL specificity in the mother. Infant CTL responses may be compromised by transmission of escape variants that arose in the mother and also those that arose in the father, if the father was the source of the mother's infection.
- ISPRTLNAW is the C consensus form of the epitope and was the autologous form in the mother, and was transmitted to her infant. By 33 weeks a new dominant form of the epitope had emerged in the infant, mSPRTLNAW, and two additional variants had arisen, one with a substitution proximal to the epitope, plSPRTLNAW, and lSPRTLNAW.

References

Pillay2005 Thillagavathie Pillay, Hua-Tang Zhang, Jan W. Drijfhout, Nicola Robinson, Helen Brown, Munira Khan, Jagadesa Moodley, Miriam Adhikari, Katja Pfafferott, Margaret E. Feeney, Anne St. John, Edward C. Holmes, Hoosen M. Coovadia, Paul Klenerman, Philip J. R. Goulder, and Rodney E. Phillips. Unique Acquisition of Cytotoxic T-Lymphocyte Escape Mutants in Infant Human Immunodeficiency

Los Alamos NATIONAL LABORATORY

Variants details

Displaying record number 53832

HXB2 Location	p24(15-23)	p24 Epitope Map
Epitope	ISPRTLNAW	Epitope Alignment
	mSPRTLNAW	escape documented in this paper
Variants	lSPRTLNAW	diminished response
	p|lSPRTLNAW	not determined
Species (MHC/HLA)	human(B57)	

Can go back to epitope entry

Variant Details

Showing all 3 variants.

Variant ID.	1413
Epitope Seq.	ISPRTLNAW
Variant Seq.	mSPRTLNAW
Mutations	I/M
Epitope Location	I1M
HXB2 Location	I15M
Mutation Type	**E**: escape documented in this paper
Method	CD8 T-cell Elispot - IFNy, Sequence
Note	This is de novo variant seen in infant by week 33 of age. The index peptide was recognized, but not the variant.

Mutation type

Note describing why the variant was designated particular mutation type

Variant ID.	1414
Epitope Seq.	ISPRTLNAW
Variant Seq.	lSPRTLNAW
Mutations	I/L
Epitope Location	I1L
HXB2 Location	I15L
Mutation Type	**DR**: diminished response
Method	CD8 T-cell Elispot - IFNy, Sequence

Los Alamos NATIONAL LABORATORY

Summary table of ~ 2800 epitope variants

CTL/CD8+ Epitope Variant Details

Download CTL/CD8+ epitope variant details as CSV or XLS files.

List of Mutation types

Data last updated at 2013-01-25 11:58:16-07

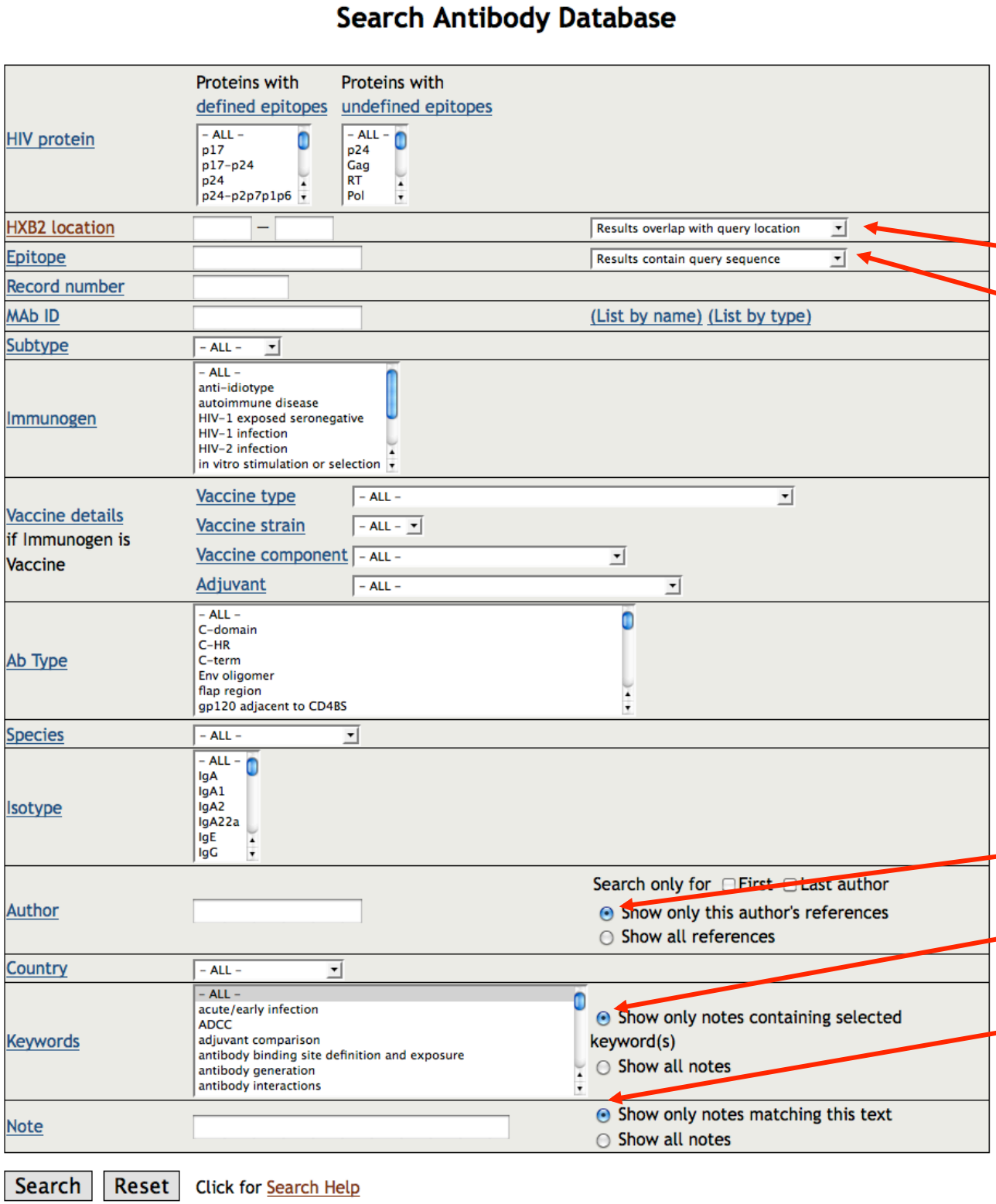
Epitope ID	Epitope Name	Variant ID	Subtype	Epitope Subtype	Variant Subtype	Protein	HXB2 start	HXB2 end	HLA	Epitope	Variant Epitope	Mutation (epitope)	Mutation (protein)	Mutation Type Code	Mutation Type Description	Methods	Note	References
54532	AI14	1016	B	B	A, M-group	p17	5	19		ASVLSGGELDRWEKI	ASVLSGGkLDaWEKI	R11A, E8K	R15A, E12K	SNSF	subtype-specific non-susceptible form	CD8 T-cell Elispot - IFNγ	No cross-recognition of this variant was seen across clades or intra-clade central sequences.	Malhotra2007
54532	AI14	1017	B	B	C	p17	5	19		ASVLSGGELDRWEKI	ASiLrGGkLDkWEKI	R11K, V3I, S5R, E8K	R15K, V7I, S9R, E12K	SNSF	subtype-specific non-susceptible form	CD8 T-cell Elispot - IFNγ	No cross-recognition of this variant was seen across clades or intra-clade central sequences.	Malhotra2007
54532	AI14	1018	B	B	B	p17	5	19		ASVLSGGELDRWEKI	ASVLSGGkLDkWEKI	R11K, E8K	R15K, E12K	SNSF	subtype-specific non-susceptible form	CD8 T-cell Elispot - IFNγ	No cross-recognition of this variant was seen across clades or intra-clade central sequences.	Malhotra2007
54532	AI14	1019	B	B	B	p17	5	19		ASVLSGGELDRWEKI	ASVLSGGELDkWEKI	R11K	R15K	SNSF	subtype-specific non-susceptible form	CD8 T-cell Elispot - IFNγ	No cross-recognition of this variant was seen across clades or intra-clade central sequences.	Malhotra2007
53591	Gag 1.2	54		B	CRF02_AG	p17	8	18		LSGGELDRWEK	LSGGkLDaWEK	E5K, R8A	E12K, R15A	SNSF	subtype-specific non-susceptible form	Intracellular cytokine staining, T-cell Elispot	CRF02 form, LSGGkLDaWEK, does not cross-react with the B clade LSGGELDRWEK elicited response.	Amara2005a
53844	GI9	1569	B			p17	11	19	B40	GELDRWEKI	GELDRWkKI	E7K	E17K	DR, LE	diminished response, literature escape	CD8 T-cell Elispot - IFNγ, Sequence	This variant from the HXB2 sequence was present in the restricting HLA-B40-carrying mother, M-1002, but was never detected in her non-HLA-B40-carrying infant, P-1031. Decreased recognition of the E17K variant relative to the index epitope was seen in the mother.	Sanchez-Merino2005
56027	GI9(p17)	1903	B	B	B	p17	11	19		GQLDRWEKI	GeLDRWEKI	Q2E	Q12E	ND	not determined	CD8 T-cell Elispot - IFNγ, Sequence	This Asian B Clade optimal epitope differs from the consensus B at one position. It is predicted to be HLA-B40 restricted. Experimentally, B clade consensus peptide was used to challenge CTL response in subjects commonly carrying the Asian B-type epitope.	Zhai2008
55632		11	A, CRF02_AG, CRF01_AE	A, AG	AE	p17	11	22		GKLDSWEKIRLR	GKLDaWEKIRLR	S5A	S15A	SSF	subtype-specific susceptible form	CD8 T-cell Elispot - IFNγ	1 subject responded to peptide GKLDSWEKIRLR from subtypes CRF02_AG and A and to peptide GKLDaWEKIRLR from subtype CRF01_AE.	Aidoo2008
54629	GAG-03	1957	B	B	C	p17	17	34		EKIRLRPGGKKKYRLKHL	EKIRLRPGGKKhYmLKHL	K12H, R14M	K28H, R30M	SSF	subtype-specific susceptible form	CD8 T-cell Elispot - IFNγ, Sequence	This Clade C consensus synthetic peptide variant from an immunodominant region, differs from the immunodominant Clade B consensus at 2 amino acids (11.1%) and both were recognized by subtype-B-infected subjects.	Zhao2007
53201	KK9	31	B			p17	18	26	A3	KIRLRPGGK	KIRLRPGGq	K9Q	K26Q	E, P	escape documented in this paper, processing	CD8 T-cell Elispot - IFNγ, Flow cytometric T-cell cytokine assay	Variant inhibits processing, resulting in rapid decline in the KK9 specific CD8+ T-cell response.	Allen2004
55770	KK9	153	B			p17	18	26	A3	KIRLRPGGK	KIRLRPGGr	K9R	K26R	SF	susceptible form	Flow cytometric T-cell cytokine assay	KIRLRPGGK was recognized by 3 patients. The autologous sequence in one patient was KIRLRPGGr which induced high frequency response.	Daucher2008
55233		790	B, CRF01_AE		B	p17	18	26	A3	KIRLRPGGK	KIRLRPGGr	K9R	K26R	IE	inferred escape	Sequence	Patient was superinfected with three strains, B1, B2 and CRF01_AE. This variant developed in B1 to include 42% of the viruses within 4 years.	Kozaczynska2007
																	This variant was seen in Donor1 and	

Los Alamos NATIONAL LABORATORY

Antibody Search

- Can search by
  - HIV protein, Epitope Sequence, Subtype, Immunogen, Vaccine Details, Species, presenting MHC/HLA, Author, Country, Keywords
  - MAb ID (Ab lists by name and by binding type are provided)
  - Ab type (by binding site, for example binding to similar region like V3 or near a common functional domain like CD4 binding site CD4Bs)
  - Isotype
- Search examples:
  - 2F5 – 1 record with 463 references
  - Ab type: gp120 CD4BS – 200 records

Antibody Search



Can search by HXB2 location,
Find overlaps, fuzzy matches
Embedded epitopes

Can show only notes and references containing selected keywords or user's text (as apposed to showing matching Ab entries with all notes)

Los Alamos NATIONAL LABORATORY

Antibody Search

Found 1 matching record:

Displaying record number 815

MAb ID	2F5 (IAM 2F5, IAM-41-2F5, IAM2F5, c2F5)	
HXB2 Location	gp160(662-667)	gp160 Epitope Map
Author Location	gp41(662-667 BH10)	
Research Contact	Hermann Katinger, Institute of Applied Microbiology, Vienna, or Polymun Scientific Inc., Vienna, Austria	
Epitope	ELDKWA	Epitope Alignment
Ab Type	gp41 adjacent to cluster II, C-term, gp41 MPER (membrane proximal external region)	
Neutralizing	L P	
Species (Isotype)	human(IgG3κ)	
Immunogen	HIV-1 infection	
Keywords	acute/early infection, adjuvant comparison, anti-idiotype, antibody binding site definition and exposure, antibody generation, antibody interactions, antibody sequence variable domain, assay development, standardization and improvement, autoantibody or autoimmunity, autologous responses, binding affinity, brain/CSF, co-receptor, complement, dendritic cells, drug resistance, enhancing activity, escape, genital and mucosal immunity, HAART, ART, HIV exposed persistently seronegative (HEPS), immunoprophylaxis, immunotherapy, immunotoxin, isotype switch, kinetics, mimics, mimotopes, mother-to-infant transmission, neutralization, rate of progression, responses in children, review, SIV, structure, subtype comparisons, supervised treatment interruptions (STI), therapeutic vaccine, vaccine antigen design, vaccine-induced immune responses, variant cross-recognition or cross-neutralization, viral fitness and reversion	

Notes

- 2F5: 2F5 neutralized infection of PBLs with various HIV-1 strains with high potency. However, 2F5 did not inhibit transcytosis of cell-free or cell-associated virus across a monolayer of epithelial cells. A mixture of 13 MAbs directed to well-defined epitopes of the HIV-1 envelope, including 2F5, did not inhibit HIV-1 transcytosis, indicating that envelope epitopes involved in neutralization are not involved in mediating HIV-1 transcytosis. When the mixture of 13 MAbs and HIV-1 was incubated with polyclonal anti-human γ chain, the transcytosis was partially inhibited, indicating that agglutination of viral particles at the apical surface of cells may be critical for HIV transcytosis inhibition by HIV-specific Abs. *Chomont2008* (**neutralization**)
- 2F5: The lipid binding properties of 2F5, and the similarity to binding properties of anti-lipid mAbs, are discussed. Potential role of liposomes containing lipid A for induction of NAbs to lipids of HIV-1 is reviewed. *Alving2008* (**autoantibody or autoimmunity**, **review**)
- 2F5: A reference panel of recently transmitted Tier 2 HIV-1 subtype B envelope viruses was developed representing a broad spectrum of genetic diversity and neutralization sensitivity. The panel includes viruses derived from male-to-male, female-to-male, and male-to-female sexual transmissions, and CCR5 as well as CXCR4 using viruses. The envelopes displayed varying degrees of neutralization sensitivity to 2F5, with 14 of 19 enevlopes sensitive to neutralization by this Ab. *Schweighardt2007* (**assay development, standardization and improvement**, **neutralization**)
- 2F5: This review summarizes data on possible vaccine targets for elicitation of neutralizing Abs and discusses whether it is more practical to design

Los Alamos NATIONAL LABORATORY

Antibody "A-list"

- List of most broadly neutralizing antibodies – currently 45 MAbs (*work in progress*)
  - Links to papers, Ab sequences and structures
  - Notes on breadth of neutralization
  - Notes on Ab contact residues
  - Notes on heavy and light chain composition

- Under "Database products" or "Epitope summary tables"

Los Alamos
NATIONAL LABORATORY

Summary of Best Neutralizing Antibodies

Download summary of best neutralizing antibodies as CSV or XLS files.

This is a list of most broadly neutralizing HIV antibodies, with links to papers, Ab sequences and structure, notes on breadth of neutralization, Ab contact or key residues and heavy and light chain composition.
Note: this is a work in progress, so not all relevant papers and antibodies are listed.

Mab	Binding site	Author Journal Pmid	First paper	Breadth of neutralization with IC50<50 µg/ml	Breadth of neutralization with IC80 or IC90<50 µg/ml	Structure, PDB ID	Ab sequence	Heavy chain	Light chain	Germline Ab sequence	Ab binding affinity	Listings of antibody contact or key residues
VRC01	CD4bs	Wu2010 Science 20616233	YES	91% of 190 isolates, representing major HIV-1 clades	86% of 190 isolates, representing major HIV-1 clades, with IC80		GI:294875838 -- heavy chain variable region GI:294875848 -- light chain variable region	V: IGHV1-02*02 D: IGHD3-16*01 (or *02) J: IGHJ1*01 or IGHJ2*01	V: IGKV3-11*01 J: IGKJ2*01	Fig. S5	Bound strongly to RSC3 and gp120 and weakly to ΔRSC3, Fig. 2 and S4.	
		Zhou2010 Science 20616231				3NGB				Fig. S12	Figs. 5, 6, S3	Env, defined by crystal structure: Fig S1. Antibody, defined by crystal structure: Fig. S9
		Wu2011 Science 21835983								Sequence, Figs. 1, S14, S18. Phylogenic analysis, Fig. 5, Fig. S13		Antibody, defined by crystal structure, compared to key residues of other CD4bs antibodies, Fig. S4.
		Scheid2011 Science 21764753		100% of 118 isolates representing major HIV-1 clades							Fig. 3, Table S9.	Antibody, defined by crystal structure in Zhou2010, Fig. 4, Fig. S3, and Fig. S4 provide comparisons with other CD4bs Nabs.
		Walker2011 Nature 21849977		93% of 162 isolates representing major HIV clades	89% of 162 isolates representing major HIV clades, with IC90							

http://hiv-dev.lanl.gov:8081/content/immunology/tables/tables.html
Epitope Summary Tables
Best Neutralizing Antibodies

Los Alamos NATIONAL LABORATORY

Tools for Immunologists

- **Sequence Locator** Finds epitope location on the reference genome
- **QuickAlign** Aligns amino acids or nucleotides against our alignments
- **PeptGen** Generates overlapping peptides for any protein
- **PepMap** Generates peptide maps in Fasta, HTML and PDF formats
- **ELF** Epitope Location Finder
- **N-Glycosite** Finds N-linked glycosylation sites
- **Mosaic** Generates candidate vaccine protein cocktails
- **Heatmaps** Displays and organizes neutralization or other quantitative data.
- And more …

HIV Molecular Immunology Database: Tools & Links

Tools Produced by the Los Alamos HIV Databases

- QuickAlign Align amino acids or nucleotides against our alignments
  - Epilign and PrimAlign have been replaced by QuickAlign
- PeptGen Generate overlapping peptides for any protein
- PepMap Generate peptide maps in Fasta, HTML and PDF formats
- Hepitope Search for hopeful epitopes based on HLA enrichment
- HLA Frequency Analysis Tools Calculate HLA frequencies or HLA linkage disequilibrium in a population
- ELF Epitope location finder
- Motif Scan Scan alignments for HLA binding motifs
  - HLA genotype/serotype dictionary
  - HLA genotype/motif dictionary
  - HLA supertype dictionaries
- HIV/SIV Sequence Locator Tool Formerly the *HXB2 Numbering Engine*
- SeqPublish Produce pretty alignments for publication
- BLAST Search sequences against our annotated HIV sequences
- ODprep/ODfit Calculate antibody titers based on concentration and optical density data
- Heatmap Display a table of numbers using colors to represent the numerical values
- Mosaic Generate candidate vaccine protein cocktails
  - Epicover Epitope coverage assessment tool
  - Posicover Positional epitope coverage assessment tool
- N-Glycosite Find N-linked glycosylation sites
- Highlighter Highlight matches and mismatches in a set of aligned sequences

- All Tools List of all software and tools in **both** the HIV sequence and immunology databases

External Tools for Epitope Prediction

- BIMAS HLA Peptide Binding Predictions Ranks potential n-mer peptides based on a predicted half-time of dissociation to HLA class I molecules
- SYFPEITHI Epitope Prediction Predicts the binding of a given amino acid sequence to a defined HLA type
- PAProC Predicts cleavages by human and yeast 20S proteasomes
- PREDEP MHC class I epitope prediction
- MHCPred Predicts MHC/peptide or TAP/peptide IC_{50} binding values
- Microsoft Research Epitope Predictor Computes the probability that a given n-mer is a T-cell epitope restricted to a given HLA allele

Los Alamos
NATIONAL LABORATORY

HIV/SIV Sequence Locator Tool

- Instantly computes position numbers of DNA or protein fragments relative to a reference strain (HXB2r for HIV-1, SMM239 for SIV)
  - Such numbers, often included in the literature, are frequently incorrect
- Shows the location of the sequence on an HIV map
- Presents protein translations of DNA sequences
- Can be used for input into the search interface, to align a new sequence you have generated with the database set
- Can also retrieve reference sequences
  - by coordinates (range of base or amino-acid positions)
  - by single position (retrieves flanking sequences)

Los Alamos
NATIONAL LABORATORY

http://www.hiv.lanl.gov/content/sequence/LOCATE/locate.html

HIV Sequence Locator Tool

Purpose: This tool has several purposes. It can find the start and end coordinates (relative to the reference strain HXB2) of your input sequence(s) and show which genes or proteins it covers, along with a graphical view of the location of your sequence(s) relative to the reference sequence. The tool will display both the nucleotide sequence and protein translation of your input as it aligns to HXB2. It will also check the reverse complement of your input sequence, and report the orientation with the best match. Another use is to retrieve a section of the HXB2 reference sequence based on its coordinates.

How to use: To find the coordinates for your sequence, either upload or paste your sequence (any format) in the box below, or (for database sequences only) enter GenBank accession numbers. To retrieve the HXB2 sequence for a set of coordinates (see HIV coordinate map), enter the coordinates and choose the region. To retrieve the entire gene or protein, enter coordinate values of "1" and "end". To retrieve a single nucleotide or range with its surrounding 42-nucleotide sequence, enter the single coordinate in the "from" field and check the box. For more details, see Sequence Locator Explanation.

Useful Links:
HXB2 numbering | SIVmm239 numbering (review articles)
HXB2 spreadsheet | SIVmm239 spreadsheet (spreadsheets with base-by-base annotation)

Find the location of a sequence

Sequence type: (•) Let program decide ( ) HIV ( ) SIV

Paste your input here [Sample Input]: SLYNTVATL

or upload your file: [ ] Browse...

-- OR --

Retrieve a region by its coordinates

Enter coordinates: from [ ] to [ ] (Enter '1' and 'end' to retrieve the entire region.)

Region: Complete

Retrieve: (•) Nucleotide or ( ) protein output

[ ] include surrounding region

Submit Reset

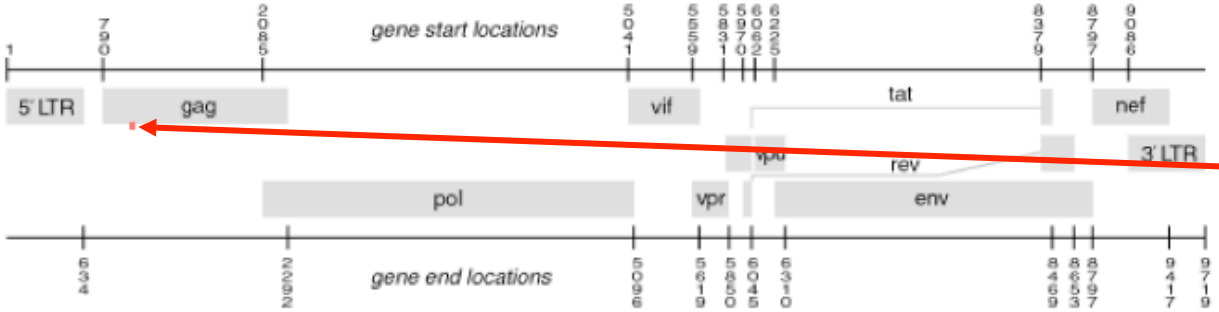
Paste or type a DNA or protein sequence here.

OR enter numeric coordiantes here.

Los Alamos NATIONAL LABORATORY

Sequence Locator: “find my sequence”

Result for SLYNTVATL Query sequence



Location in genome mapped in red.

Table of protein regions touched by query sequence. AA = amino acid, NA = nucleic acid.					
CDS	**AA** position relative to protein start in HXB2	**AA** position relative to query sequence start	**AA** position relative to polyprotein start in HXB2	**NA** position relative to CDS start in HXB2	**NA** position relative to HXB2 genome start
Gag	77 → 85	1 → 9	NA	229 → 255	1018 → 1044
p17	77 → 85	1 → 9	NA	229 → 255	1018 → 1044

Numeric coordinates useful for entry on search form

Alignment of the query sequence to HXB2 (Similarity 100.0%):

```
Query SLYNTVATL  9
      :::::::::
 HXB2 SLYNTVATL
```

DNA and protein sequence displayed

Los Alamos NATIONAL LABORATORY

Sequence Locator: "Retrieve from coordinates"

-- OR --

Retrieve a region by its coordinates

Enter coordinates: from 77 to 85 (Enter '1' and 'end' to retrieve the entire region.)

Region: Gag

Retrieve: ○ Nucleotide or ◉ protein output

☐ include surrounding region

Submit Reset

Include surrounding region

Reference Strain	Type	Region	Start	End
HXB2	pro	complete	77	85

Retrieved Sequence:

SLYNTVATL

Reference Strain	Type	Region	Start	End
HXB2	pro	complete	56	106

Retrieved Sequence:

GCRQILGQLQPSLQTGSEELRSLYNTVATLYCVHQRIEIKDTKEALDKIEE

50 aa long stretch

Los Alamos NATIONAL LABORATORY

QuickAlign

- Generates an alignment of your HIV-1 amino acid or nucleotide sequence against our web alignments
- Can be used to align epitopes, functional domains, or any protein or nucleotide region of interest
- Calculates frequency of variants to the query sequence and summarizes both by subtype and all subtypes together
- Calculates frequency of amino acid or nucleotide by position and summarizes both by subtype and all subtypes together

Los Alamos
NATIONAL LABORATORY

QuickAlign

formerly Epilign and Primalign

Purpose: Align a desired region from our Web alignments, with or without user-provided sequence(s). Details below.

Retrieve alignment(s) based on sequence

Paste your sequence(s) here [Sample Input]

or upload sequence file Browse...

-- OR leave both fields above blank, and --

Retrieve alignment(s) based on coordinates

Sequence coordinates 1 start end end

Gene/region/protein Complete

Options

Organism HIV1 HIV2 SIV

Sequence type nucleotide protein let program decide

Alignment type to use 'Web' alignment (all complete sequences)

Delete Gaps and shift sequence toward C-terminus (protein only) yes no

Display wide output yes no

Calculate frequency by position cut-off 95 %

Include surrounding region

Submit Reset

os Alamos
TIONAL LABORATORY

QuickAlign: example of output

- Query peptide: SLYNTVATL
- Sequence names include subtype, country and year of sampling
- Identical sequences are shown in red

Query:	SLYNTVATL
Query Length:	9
HXB2 Location:	Gag 77-85 = p17 77-85
Alignment:	GAG, 458 sequences

Summarize

```
Query                SLYNTVATL
A1.KE.86.ML170       --F------
A1.KE.94.Q23         --F------
A1.SE.94.SE7253      --F----V-
A1.SE.94.SE7535      ---------
A1.SE.95.SE8538      ---------
A1.SE.95.SE8891      ---------
A1.SE.95.UGSE8131    ---------
A1.TZ.97.97TZ03      --F----V-
```

Los Alamos NATIONAL LABORATORY

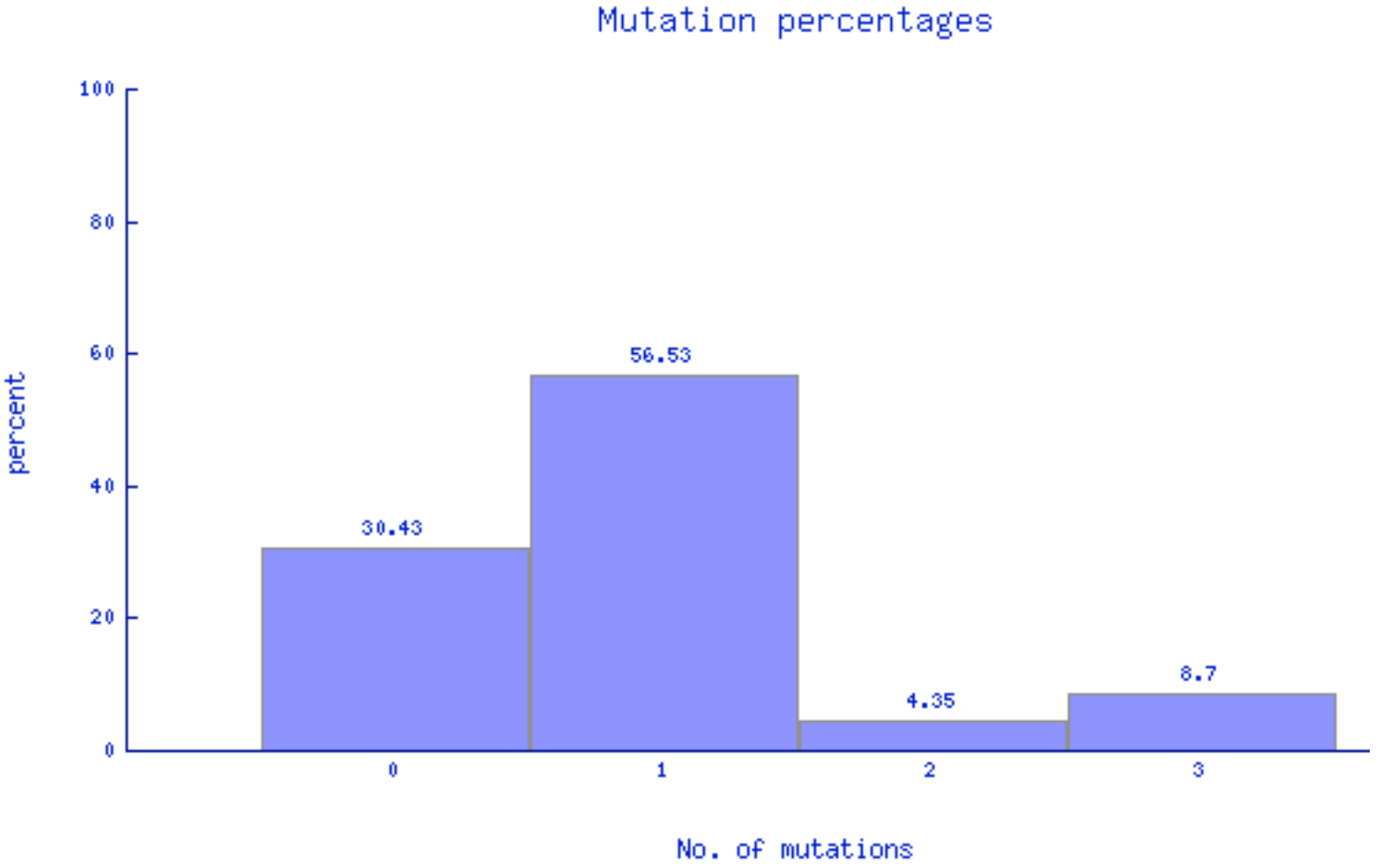
QuickAlign: sequence variant summary

Summary for subtype A

Variant	Count	Percent
SLYNTVATL		
--F------	11	47.83
---------	7	30.43
--F--I-V-	1	4.35
--F----V-	1	4.35
-------V-	1	4.35
----L----	1	4.35
--F-A--V-	1	4.35

Total sequences = 23

Number of variants = 7



- Variant frequency summary by subtype and all subtype together

Los Alamos NATIONAL LABORATORY

QuickAlign: Frequency by position

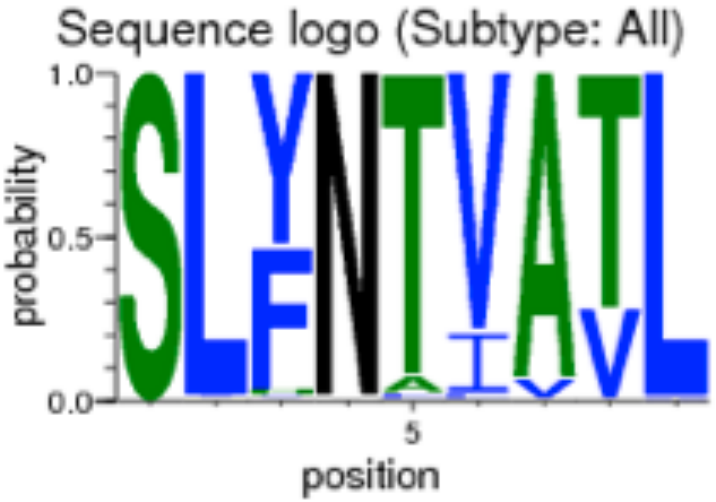
Frequency by position

Go to top

See full raw counts

cutoff: 95%

Position	Percentage and raw count of non-gap	Non-gap/total (percentage)
1	S: 99.90% (3113) other: 0.10% (3)	3116/3119 (100.00%)
2	L: 98.90% (3068) other: 1.10% (34)	3102/3119 (99.55%)
3	Y: 52.71% (1633) F: 43.77% (1356) other: 3.52% (109)	3098/3119 (99.42%)
4	N: 99.68% (3104) other: 0.32% (10)	3114/3119 (99.94%)
5	T: 92.86% (2887) A: 5.05% (157) other: 2.09% (65)	3109/3119 (99.78%)
6	V: 79.35% (2448) I: 18.15% (560) other: 2.50% (77)	3085/3119 (99.01%)
7	A: 92.95% (2889) V: 6.53% (203) other: 0.51% (16)	3108/3119 (99.74%)
8	T: 72.52% (2254) V: 27.06% (841) other: 0.42% (13)	3108/3119 (99.74%)
9	L: 99.00% (3078) other: 1.00% (31)	3109/3119 (99.78%)



Los Alamos NATIONAL LABORATORY

PepMap

- Maps an input set of peptides on HXB2 reference sequence
- Maps peptide set on HXB2 reference sequences
- Can be used to map epitopes, functional domains, or any protein region of interest
- Peptide name can contain any kind of useful information, like peptide number, HLA, reactive or not reactive, autologous sequence, patient ID etc
- Creates maps of overlapping peptides on proteins to aid in peptide design for mapping epitopes, both in HTML and fasta formats

PepMap

Purpose: The PepMap tool maps an input set of peptides on the HIV reference sequence HXB2. Peptide maps are generated in Fasta, PDF and HTML formats.

Details: This tool can be used to map epitopes, functional domains, or any protein region of interest. In addition to generating peptide maps in Fasta, HTML and PDF formats, the tool displays a table with peptide locations both in HTML and text formats. For details, see PepMap Explanation.

Input

Paste or upload your input peptide set

```
p17_19_28       IRLRPGGKKK
p17_86_101      YCVHQRIEIKDTKEAL
p24_21_40       NAWVKVVEEKAFSPEVIPMF
p24_46_59       GATPQDLNTMLNTV
RT_14_23        PGMDGPKVKQ
RT_18_26        GPKVKQWPL
Nef_37_45       LEKHGAITS
```

or upload your file [Browse...]

Options

Generate Maps of the Alignment [x]

Submit Reset

Please be patient. Depending on the size of your input file, the speed of your internet connection, and the load on our server, it can take several minutes. Do not resubmit your sequences; you won't get a result any faster - you'll just load up our server and make the process slower.

last modified: Thu Sep 20 10:45 2012

PepMap

Input:

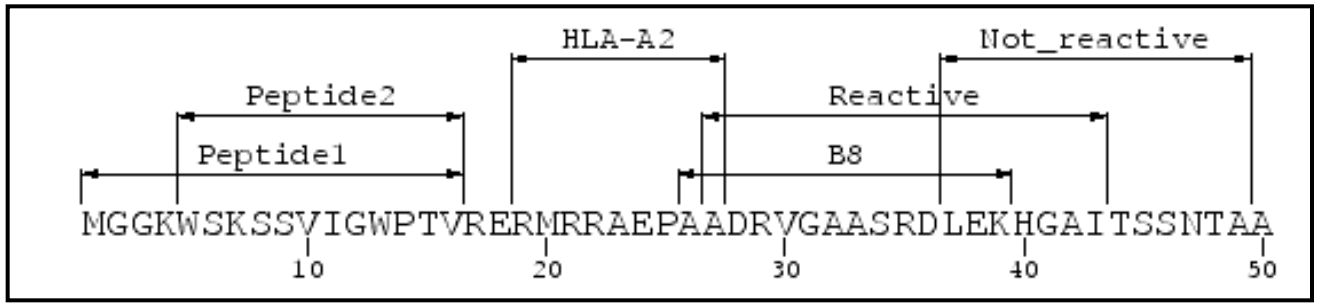
Peptide1	MGGKWSASSVIGGPTV
Peptide2	WSKSSVIGWVTV
HLA-A2	RMRRAEPAV
B8	AADRVGAASRDLEK
Reactive	ADRVGAASRDLEKHGAI
Not_reactive	LEKHGAITSSNTA

PepMap output

Location table

Epitope Name	Query Peptide	Reference Peptide	Protein	AA position in Protein	Polyprotein	AA position in Polyprotein	Similarity%
Peptide1	MGGKWSASSVIGGPTV	MGGKWSKSSVIGWPTV	Nef	1-16	-	-	87.5
Peptide2	WSKSSVIGWVTV	WSKSSVIGWPTV	Nef	5-16	-	-	91.7
HLA-A2	RMRRAEPAV	RMRRAEPAA	Nef	19-27	-	-	88.9
B8	AADRVGAASRDLEK	AADRVGAASRDLEK	Nef	26-39	-	-	100.0
Reactive	ADRVGAASRDLEKHGAI	ADRVGAASRDLEKHGAI	Nef	27-43	-	-	100.0
Not_reactive	LEKHGAITSSNTA	LEKHGAITSSNTA	Nef	37-49	-	-	100.0

Peptide map on HXB2 (PDF)
Can be used for presentations



Peptides aligned to HXB2 (FASTA)
Can be used for further analysis

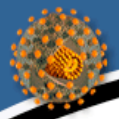
```
>B.FR.83.HXB2_LAI_IIIB_BRU_K03455
MGGKWSKSSVIGWPTVRERMRRAEPAADRVGAASRDLEKHGAITSSNTAA

>Peptide1
MGGKWSASSVIGGPTV--------------------------------
>Peptide2
----WSKSSVIGWVTV--------------------------------
>HLA-A2
----------------RMRRAEPAV----------------------
>B8
------------------------AADRVGAASRDLEK----------
>Reactive
-------------------------ADRVGAASRDLEKHGAI-------
>Not_reactive
---------------------------------LEKHGAITSSNTA-
```

Los Alamos NATIONAL LABORATORY

PeptGen

- Creates maps of overlapping peptides on proteins to aid in peptide design for mapping epitopes
- Consensus sequences for all HIV subtypes for all proteins are available
- Use alignments to design comparable sets of peptides (for example, to compare clades)
- INPUT
  - Query sequence or aligned sequences
  - Desired length of peptides, peptide overlap,
  - Forbidden C- and N-terminal amino acids
- OUTPUT
  - Maps of overlapping peptides (forbidden AAs are taken into account)
  - Simplified output for ordering
  - Highlighted forbidden amino acids
  - Hydrophobicity scores for the peptides are available



HIV sequence database

DATABASES | SEARCH | ALIGNMENTS | TOOLS | PUBLICATIONS | GUIDES | Search Site

PeptGen Peptide Generator

Purpose: Given an amino acid sequence, this tool generates and displays sets of overlapping peptides that can be used for peptide design and epitope mapping.

How to use: Paste amino acid sequence(s) in any valid format into the window below, or upload a file of sequences. Change the default values for the other parameters as needed and hit Submit. You can also use an HIV consensus sequence as input.

For additional details, see PeptGen Explanation.

Input

Paste your input here
[SAMPLE INPUT - single sequence]
[SAMPLE INPUT - alignment]

or upload your file — Browse...

If you are submitting an alignment check this box ☐

Options

Make peptides of length	18	C-terminal forbidden amino acids	
Overlap peptides by	10	N-terminal forbidden amino acids	
Shorten by	3	Apply proline rule	◉ Yes ○ No
Lengthen by	2	Calculate hydropathy	◉ Yes ○ No
		Color amino acids	◉ Yes ○ No

Output

Produce simple output ☐

Duplicate peptides ◉ flag ○ remove

Gaps ◉ remove ○ save

Output style ◉ classic ○ new

Submit Reset

Los Alamos NATIONAL LABORATORY

PeptGen: output

Classic output:

PeptGen Results

Download a copy of these results in format: Text PostScript PDF

```
                  Word length: 18
Overlap consecutive peptides by: 10
                   Shorten by:
                  Lengthen by:
   Forbidden C-term amino acids:
   Forbidden N-term amino acids:
   Number of peptides generated: 9
               Sequence names: CON_B
                               CON_C
                               CON_G
```

```
HIVWASRELERFAVNPGLLETSEGCRQILGQLQPSLQTGSEELRSLYNTVATLYCVHQRIEVKDTKEALEKIEEEQNKSK
HIVWASRELERFAVNPGL (18)
-L-----------L----
-L-----------L--D-
        LERFAVNPGLLETSEGCR (18)
        -----L-----------K
        -----L--D----A---Q
              GLLETSEGCRQILGQLQP (18)
              ---------K--IK----
              D----A---Q--M-----
                     CRQILGQLQPSLQTGSEE (18)
                     -K--IK----A----T--
                     -Q--M-----A----T--
                            QPSLQTGSEELRSLYNTV (18)
                            --A----T----------
                            --A----T------F---
                                   EELRSLYNTVATLYCVHQ (18)
                                   ----------------E
                                   ------F-----------
                                          TVATLYCVHQRIEVKDTK (18)
                                          ---------EK---R---
                                          ------------------
                                                 HQRIEVKDTKEALEKIEE (18)
                                                 -EK---R------D----
                                                 --------------EV-K
                                                        TKEALEKIEEEQNKSK (16)
                                                        -----D---------Q
                                                        ------EV-KI-K--Q
```

Distinct

Duplicate

Simple output, “new”

```
13 QPSLQTGSEELRSLYNTV 5_1.1
14 QPALQTGTEELRSLYNTV 5_2.1
15 QPALQTGTEELRSLFNTV 5_3.1

16 EELRSLYNTVATLYCVHQ 6_1.1
17 EELRSLYNTVATLYCVHE 6_2.1
18 EELRSLFNTVATLYCVHQ 6_3.1

19 TVATLYCVHQRIEVKDTK 7_1&3.2
20 TVATLYCVHEKIEVRDTK 7_2.1

21 HQRIEVKDTKEALEKIEE 8_1.1
22 HEKIEVRDTKEALDKIEE 8_2.1
23 HQRIEVKDTKEALEEVEK 8_3.1

24 TKEALEKIEEEQNKSK 9_1.1
25 TKEALDKIEEEQNKSQ 9_2.1
26 TKEALEEVEKIQKKSQ 9_3.1
```

Los Alamos NATIONAL LABORATORY

N-Glycosite

- Highlights and tallies predicted N-linked glycosylation sites (Nx[ST] patterns, where x can be any amino acid)
- NP[ST] pattern can be excluded

```
                                    210        220        230        240        250        260        270
                                    |          |          |          |          |          |          |
A1.KE.93.Q23-17                     QACPKVSFEP IPIHYCTPAG FAILKCKDEG FNGTGL--CK NVSTVQCTHG IKPVVSTQLL LNGSLAEKNI
B.FR.HXB2                           QACPKVSFEP IPIHYCAPAG FAILKCNNKT FNGTGP--CT NVSTVQCTHG IRPVVSTQLL LNGSLAEEEV
C.BR.92.92BR025                     QACPKVSFDP IPIHYCAPAG YAILKCNNKT FNGTGP--CN NVSTIQCTHG TKPVVSTQLL LNGSLAEEEI
D.UG.94.94UG1141                    QACPKMTFEP IPIHYCAPAG FAILKCNEKK FNGTGP--CK NVSTVQCTHG IKPVVSTQLL LNGSLAEEEI
01_AE.CF.90.90CF11697               QACPKVTFDP IPIHYCTPAG YAILKCNEKN FNGTGP--CK NVSSVQCTHG IKPVVSTQLL LNGSLAEEDI
02_AG.CM.97.97CM-MP807              QACPKVSFEP IPIHFCAPAG FAILKCKDKE FNGTGP--CK NVSTVQCTHG IKPVVSTQLL LNGSLAEEKV
CPZ.CM.-.CAM3                       QACPKTSFEP IPIHYCATPG YAIMKCNMPN FNGTGTGRCN NISTVQCTHG IRPVVTTQLI LNGSVAENKT
O.CM.-.ANT70                        QACPKVSFEP IPIHYCAPAG YAIFKCNSTE FNGTGT--CR NITVVTCTHG IRPTVSTQLI LNGTLSKGKI
```

N-glycosylation Sites In Each Submitted Suquence

Sequence Name	N-glycosylation Sites Numbers
A1.KE.93.Q23-17	23
B.FR.HXB2	24
C.BR.92.92BR025	23
D.UG.94.94UG1141	24
01_AE.CF.90.90CF11697	24
02_AG.CM.97.97CM-MP807	24
CPZ.CM.-.CAM3	25
O.CM.-.ANT70	25

A total of 192 N-glycosylation sites in 8 sequences have been found.

The total N-Linked glycosylation sites is 192. (Click the numbers and see the details)

- View the N-Linked glycosylation sites in NX[ST]pattern.
  - The single N-linked glycosylation site count is: 192.
    - The NXS combination count is: 62.
    - The NXT combination count is: 130.
  - The contiguous N-linked glycosylation site (NN[ST][ST]) count is: 0.

Los Alamos NATIONAL LABORATORY

N-Glycosite (continuation)

- Tallies number of N-linked glycosylation sites per alignment position and displays as a downloadable table

Number of N-glycosylation Sites by Position

Download tab-delimited file

Position	31	60	69	102	105	109	110	111	112	115	122	128	145	149	174	179	182	184	185	195	227
A1.KE.93.Q23-17		N			N				N				N	N						N	
B.FR.HXB2		N					N				N		N	N		N				N	
C.BR.92.92BR025		N			N						N		N	N		N		N		N	
D.UG.94.94UG1141		N		N							N		N	N		N			N	N	
01_AE.CF.90.90CF11697		N						N			N	N	N	N		N			N	N	
02_AG.CM.97.97CM-MP807		N				N				N	N		N	N					N	N	
CPZ.CM.-.CAM3		N	N					N			N		N	N		N				N	
O.CM.-.ANT70	N	N												N	N		N			N	N
Total	**1**	**8**	**1**	**1**	**2**	**1**	**1**	**2**	**1**	**1**	**6**	**1**	**7**	**8**	**1**	**5**	**1**	**1**	**3**	**8**	**1**

Los Alamos NATIONAL LABORATORY

ELF

- If you have a peptide that reacts with CD8+ T cells from a person with known HLA type, enter:
  - The peptide that reacts with CD8+ T-cells
  - The HLA type of the person with the reactive CD8+ T cells

- ELF will help identify the possibly reactive epitope by
  - Highlighting appropriate HLA anchor motifs in the peptide
  - Aligning all known epitopes embedded in the peptide from the database to your query sequence, with links to epitope entries
  - Finding potential epitopes based on Immune Epitope Database (IEDB) binding predictions http://www.immuneepitope.org/
- Other useful information provided:
  - Genomic location of your peptide
  - Database records for known CTL epitopes in this region, regardless of HLA.

Los Alamos
NATIONAL LABORATORY

ELF

Epitope Location Finder

Purpose: search a submitted protein sequence for (1) known epitopes from our immunology databases, (2) epitopes predicted by consensus binding motifs, and (3) epitopes predicted by the IEDB binding algorithm. For details see ELF Explanation.

Input

Paste protein sequence: DTVLEDMNLPGRWKPKMIG <50 amino acids, raw format

Options

Show known epitopes: ☑ from CTL and Helper databases

Find potential epitopes: ☑ based on anchor residues

Choose HLA(s) (Class I and Class II)

Use control-click for multiple selection

By genotype	By serotype
A*3004	A33(19)
A*3101	A69(28)
A*3201	A68(28)
A*3303	A30(19)
A*6601	A66(10)
A*6801	A1
A*6802	A2

Find potential epitopes: ☑ based on IEDB binding predictions

Choose HLA(s) or MHC(s) (synchronized with genotype selections above)

HLA Class I	HLA Class II
A*6611	DRB3*0224
A*6612	DRB3*0225
A*6613	DRB3*0301
A*6614	DRB3*0303
A*6615	DRB4*0101
A*6801	DRB4*0103
A*6802	DRB5*0101

Animal MHC Class I	Animal MHC Class II
chimpanzee	***mouse***
Patr-A*0101	H2-IAb
Patr-A*0201	H2-IAd
Patr-A*0301	H2-IEd
Patr-A*0302	
Patr-A*0401	
Patr-A*0402	

Display binders: ◉ Show 1 best binder(s) per MHC

○ Show below 20 percentile rank (1-100) per MHC

E-mail result: ☐ Predictions are slow. For more than a few HLAs/MHCs, we recommend e-mailed result.

Submit Reset

HLA selection is synchronized between 2 analysis options

You can choose how many top binders to show per MHC, or use a binding percentile rank cutoff

Los Alamos NATIONAL LABORATORY

ELF results 1:

Epitopes from our CTL database aligned to your query sequence

Bold red letters indicate residues that differ from the query sequence. The symbol ➤ means the HLA of the epitope matches one of your submitted HLAs. Click on the epitope to see full database entry. Click on "align" to align the epitope to the sequence database via QuickAlign.

Epitopes shown here are completely within the bounds of your query. Epitopes that overlap the ends of your query are included in the "View database records" links above.

Download this alignment in format table

DTVLEDMNLPGRWKPKMIG

DTVLEEMNL A*6802 align ➤

DTVLEDINL A*6802 align ➤

DTVLEEWNL A*6802 align ➤

DTVLEEMNL A68 align

DTVLEEMNL A28 align

DTVLEDMNL align

EEMNLPGRW B44 align

EEINLPGKW B44 align

EEMNLPGRW B*4402 align

EEMNLPGRW B*4403 align

EEMNLPGRW B18,B40,B44 align

EDMNLPGRW align

EEMNLPGRW B*44 align

EEINLPGKW B*4403 align

EEMNLPGRW align

LPGRWKPKMI Cw3 align

LPGRWKPKMI B7 align

Clicking on an epitope takes you to respective CTL or Helper epitope Database entries

Los Alamos NATIONAL LABORATORY

ELF results 2:

Potential epitopes based on anchor residues

These peptides have C-terminal anchor residues, highlighted in **blue**, and internal anchors highlighted in magenta. These anchor residues match one or more motifs associated with the submitted HLA.

Download this alignment in format table

```
DTVLEDMNLPGRWKPKMIG
DTVLEDMNL (A*0205 ........[L])
DTVLEDMNL (A*6802 .[TV]......[VL])
 TVLEDMNLP (A*0206 .[VQ].......)
   LEDMNLPGR (DRB5*0101,DRB5*0101 [FYLM]..[QVIM]....[RK])
```

Los Alamos NATIONAL LABORATORY

ELF results 3:
Potential epitopes based on IEDB database MHC binding predictions, by Alexander Sette's group

Potential epitopes based on IEDB binding predictions

Top binders for each MHC are highlighted in **blue**.
Prediction method: IEDB recommended
Low percentile = good binders
Show up to 1 binder(s) per MHC

Class I

Selected allele(s): A*6802, B*1501

Download this alignment in format table

```
DTVLEDMNLPGRWKPKMIG (Click MHC to see full list of IEDB predictions for that MHC)
      DMNLPGRW       B*1501 (26)
       MNLPGRWK      A*6802 (3.0)
```

Clicking on MHC links to the full list of IEDB predictions for that MHC (see next slide)

Class II

Selected allele(s): DRB5*0101

Download this alignment in format table

```
DTVLEDMNLPGRWKPKMIG (Click MHC to see full list of IEDB predictions for that MHC)
 TVLEDMNLPGRWKPK     DRB5*0101 (17.17)
```

Los Alamos NATIONAL LABORATORY

ELF results 3:

Potential epitopes based on IEDB database MHC binding predictions, by Alexander Sette's group

IEDB Analysis Resource

Home | Help | Example | Reference | Download | Contact

MHC-I binding predictions - Prediction Results

Input Sequences

#	Name	Sequence
1	sequence 1	DTVLEDMNLPGRWKPKMIG

Prediction method: IEDB recommended | Low percentile = good binders

Check to expanded the result: ☐

Allele	#	Start	End	Peptide Length	Sequence	Method used	Percentile Rank
HLA-B*15:01	1	6	13	8	DMNLPGRW	*NetMHCpan*	26
HLA-B*15:01	1	3	13	11	VLEDMNLPGRW	*NetMHCpan*	27
HLA-B*15:01	1	3	11	9	VLEDMNLPG	Consensus (*ANN,SMM,CombLib_Sidney2008*)	27.60
HLA-B*15:01	1	8	17	10	NLPGRWKPKM	*NetMHCpan*	31
HLA-B*15:01	1	7	17	11	MNLPGRWKPKM	*NetMHCpan*	35
HLA-B*15:01	1	2	9	8	TVLEDMNL	*NetMHCpan*	36
HLA-B*15:01	1	2	11	10	TVLEDMNLPG	*NetMHCpan*	47
HLA-B*15:01	1	4	11	8	LEDMNLPG	*NetMHCpan*	48

Alamos
IAL LABORATORY

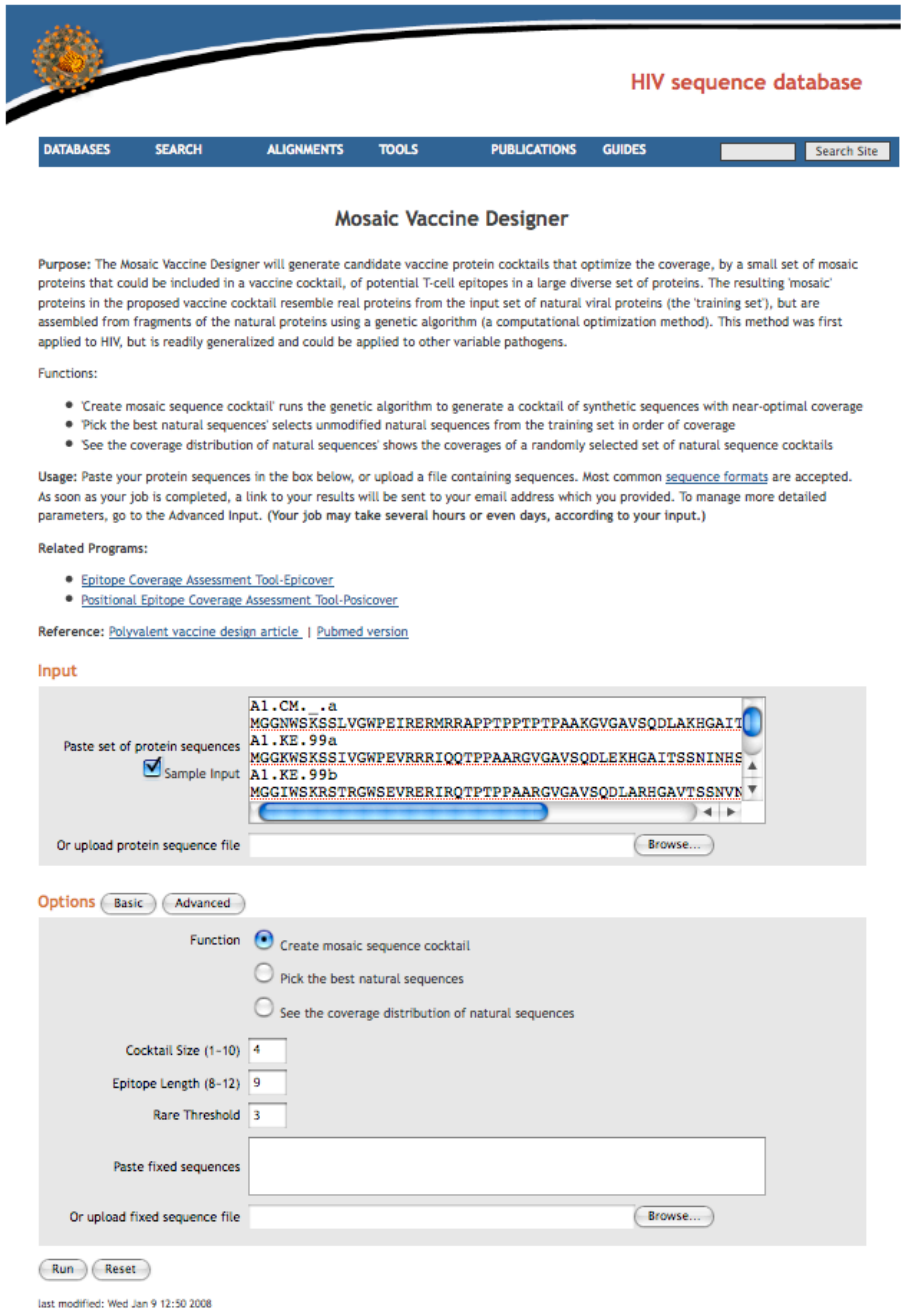
Mosaic vaccine tools

Mosaic Vaccine Designer: The Mosaic Vaccine Designer will generate candidate vaccine protein 'cocktails' that optimize coverage of potential T-cell epitopes found in a given background set of protein sequences.

Epitope Coverage Assessment: Alignment independent “n-mer” coverage of sequences by vaccines or peptides.

Positional Epitope Coverage Assessment: Alignment dependent coverage of sequences by vaccines or peptides.

Mosaic Vaccine Designer



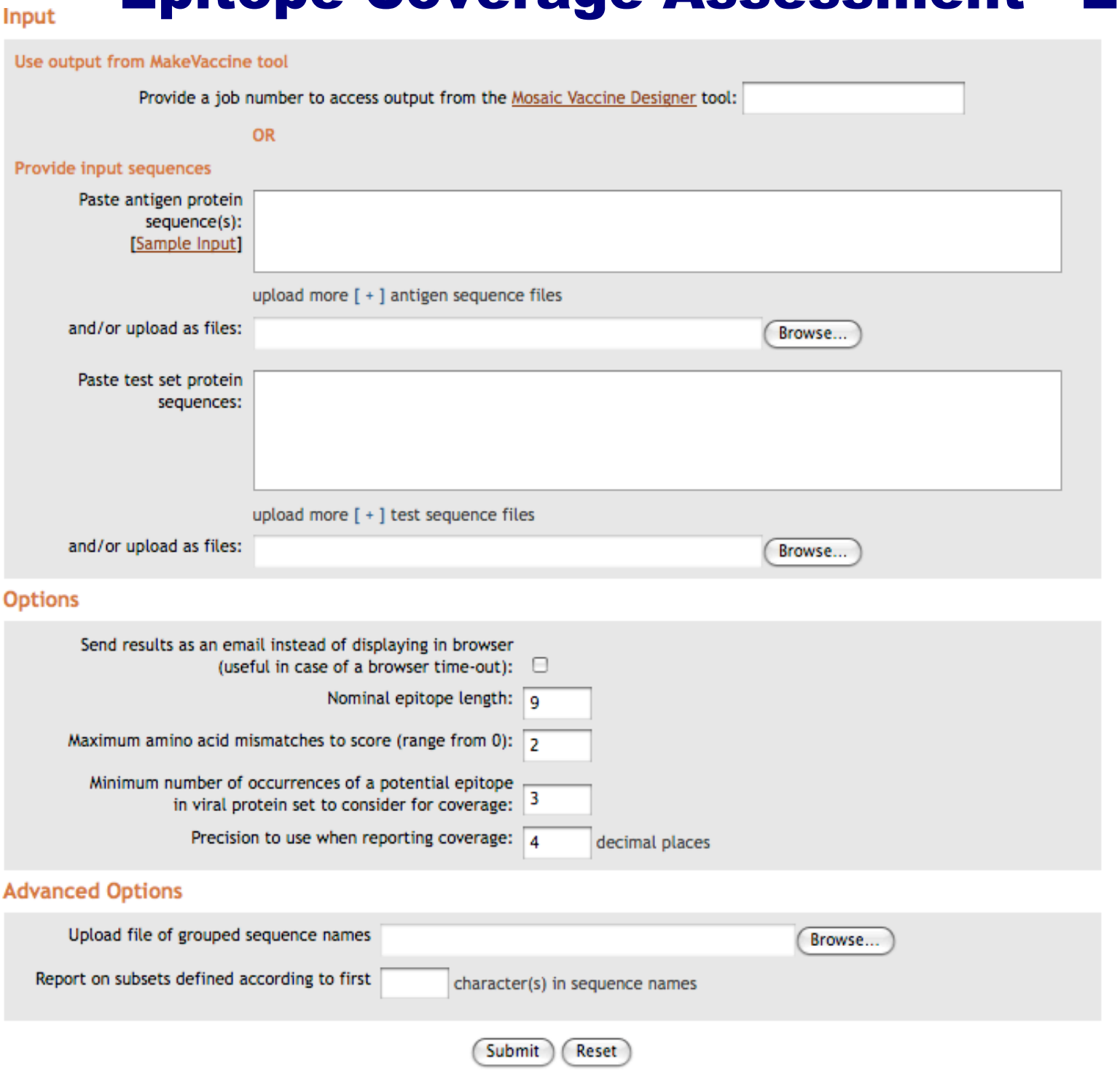
Input: protein sequence set for a target population, does not need to be aligned.

Number of mosaic proteins in the set.

Epitope length.

Los Alamos NATIONAL LABORATORY

Epitope Coverage Assessment - Epicover



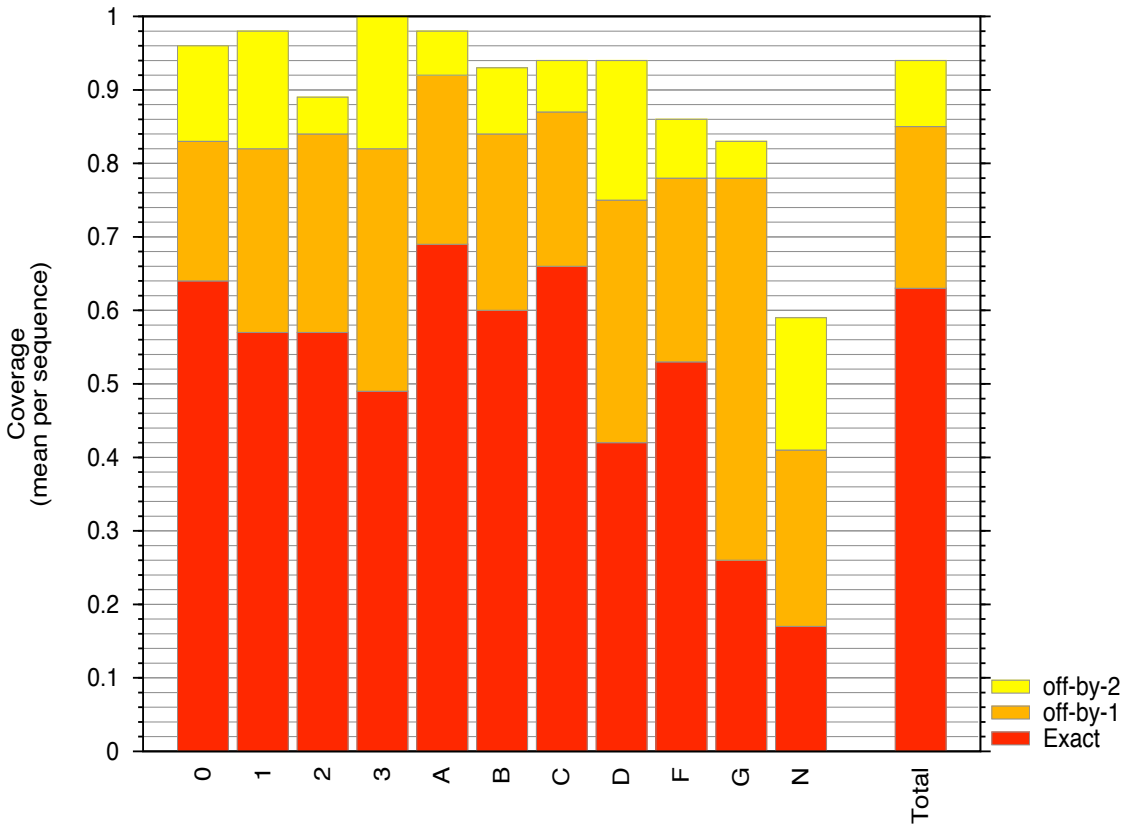
Input:
Vaccine set
Test set

Can report on subsets defined according to the first several characters in sequence names or user-defined subsets

Los Alamos NATIONAL LABORATORY

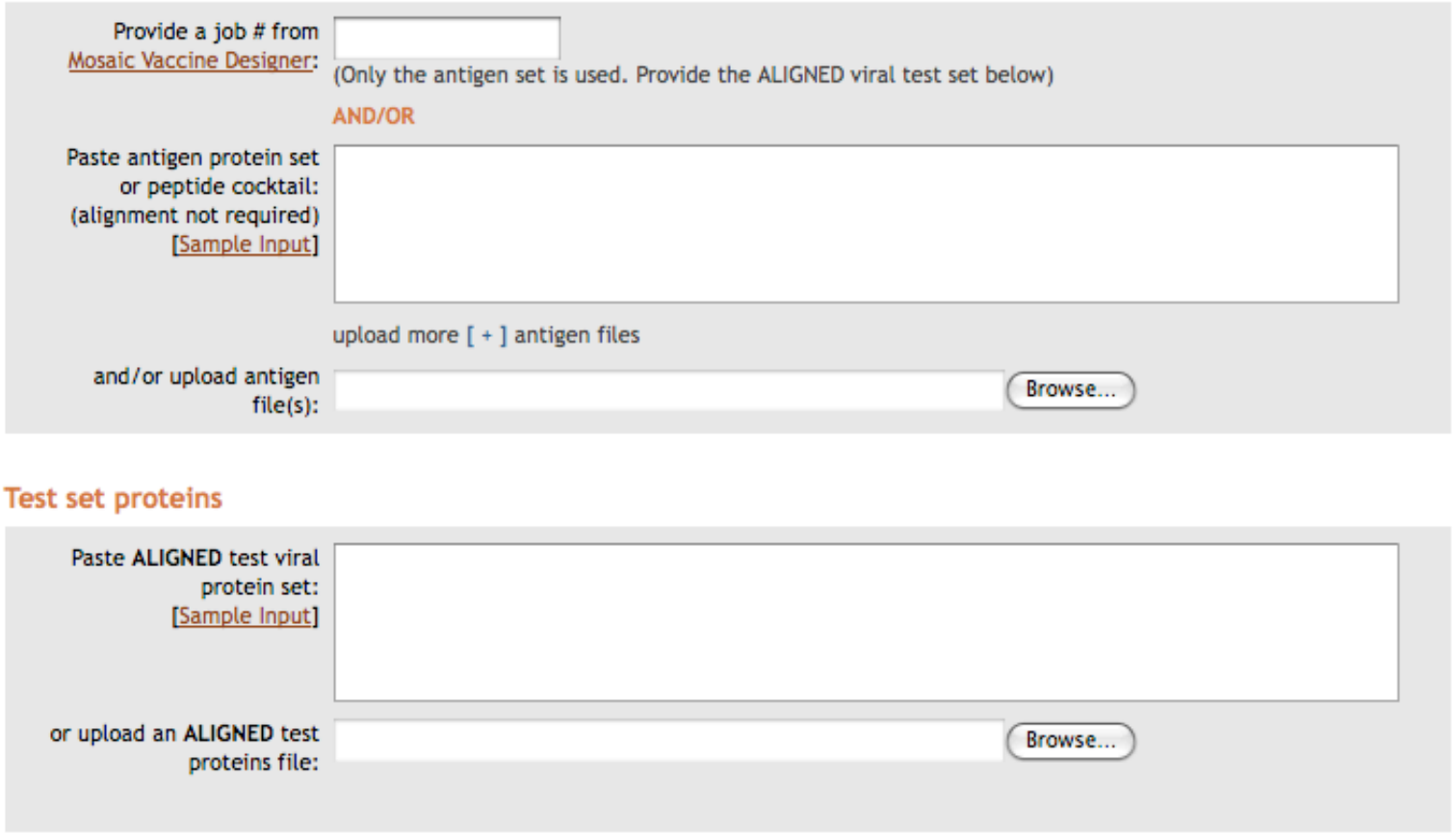
Epicover output

vaccine set	subset	subset count	Off-by-0	Off-by-1	Off-by-2	rare(<3,>1)	unique	absent
vaccine_set_from_user	**Total**	**63**	**0.6615**	**0.8914**	**0.9660**	**104**	**114**	**334**
vaccine_set_from_user	A	11	0.7232	0.9429	0.9935	47	36	334
vaccine_set_from_user	B	11	0.6378	0.8845	0.9755	25	19	334
vaccine_set_from_user	C	35	0.6921	0.8994	0.9637	51	45	334
vaccine_set_from_user	D	4	0.4217	0.7546	0.9443	4	9	334
vaccine_set_from_user	F	1	0.5300	0.7800	0.8600	4	5	334
vaccine_set_from_user	G	1	0.2597	0.7792	0.8312	0	0	334



Los Alamos
NATIONAL LABORATORY

Positional Epitope Coverage Assessment - Posicover



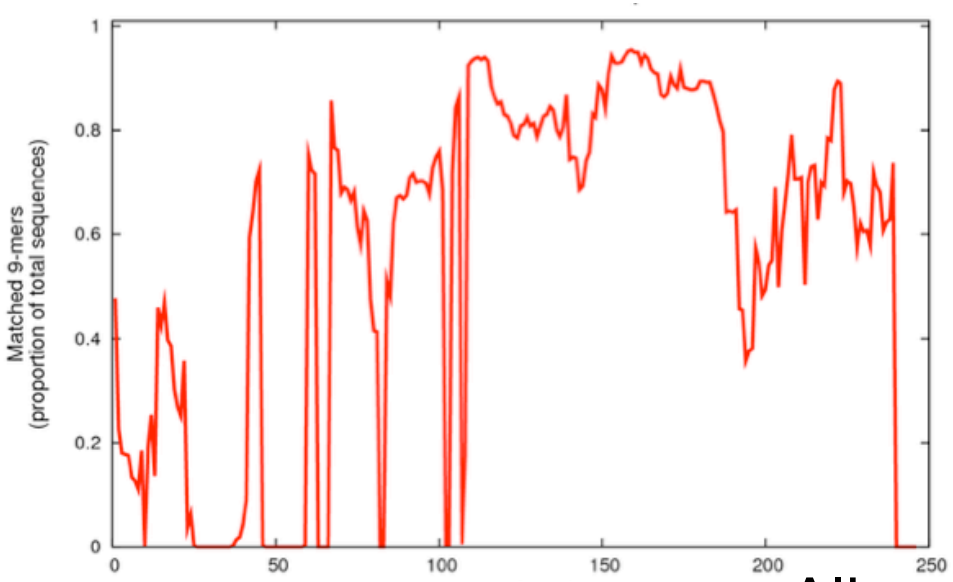
Input:

Vaccine set ALIGNED test set

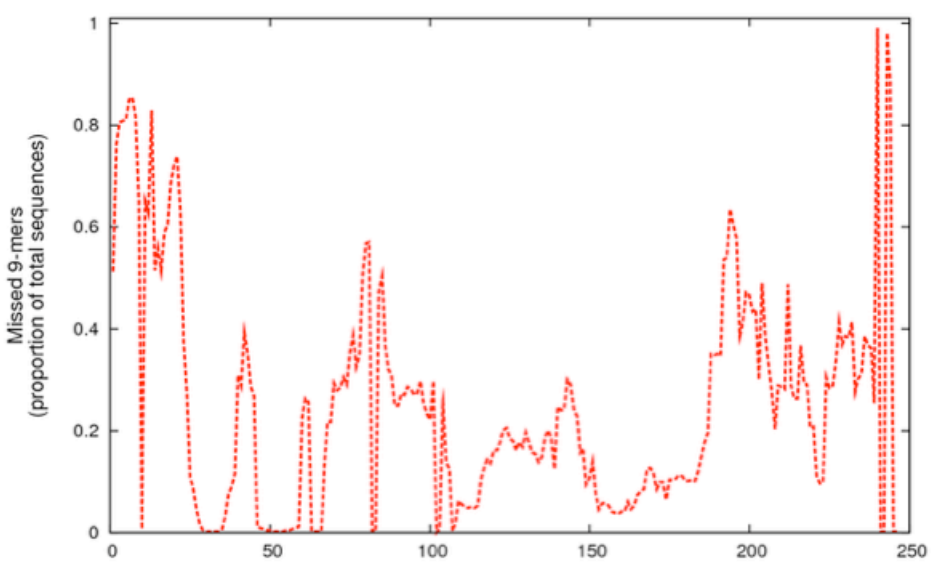
Los Alamos NATIONAL LABORATORY

Examples of Posicover outputs

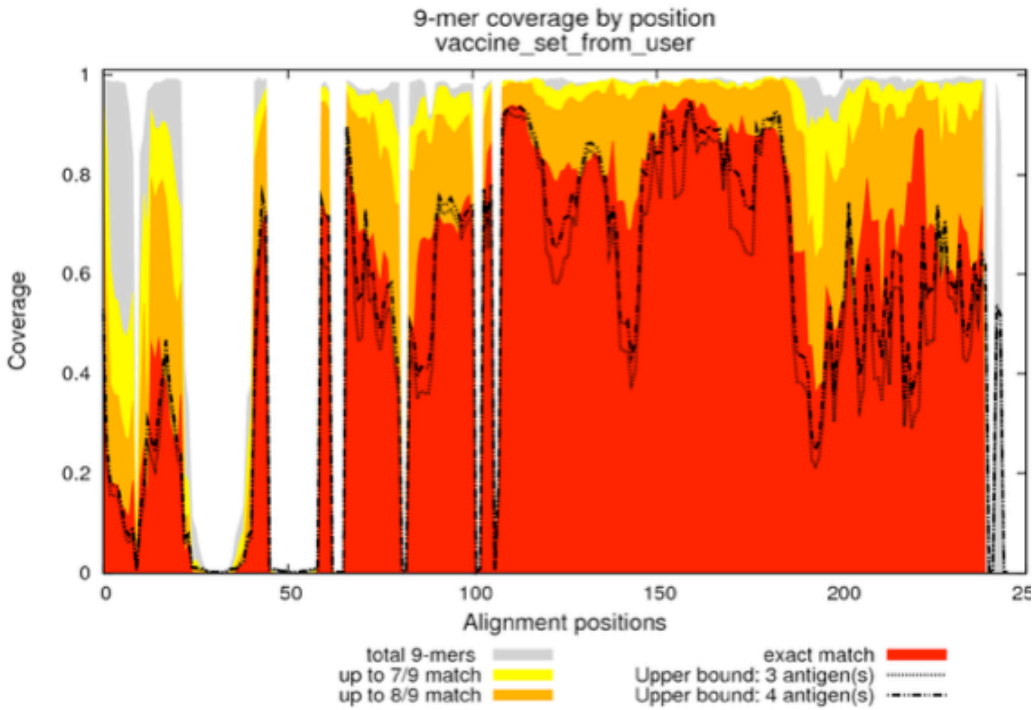
Matched 9-mers

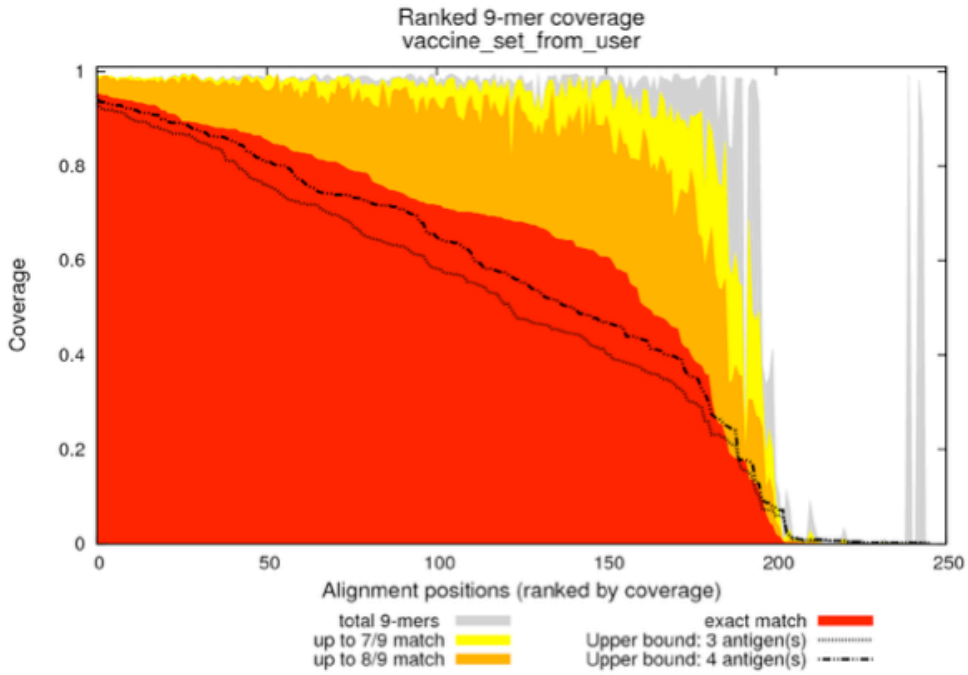


Missed 9-mers



Alignment positions





Examples of Posicover outputs

User's sequence alignment:

Each aa is represented as a single-colored square

Los Alamos
NATIONAL LABORATORY

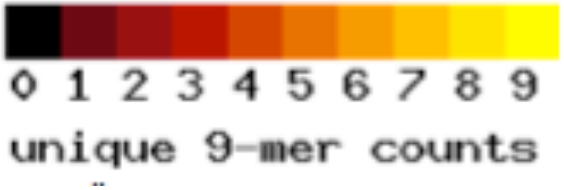
Examples of Posicover outputs

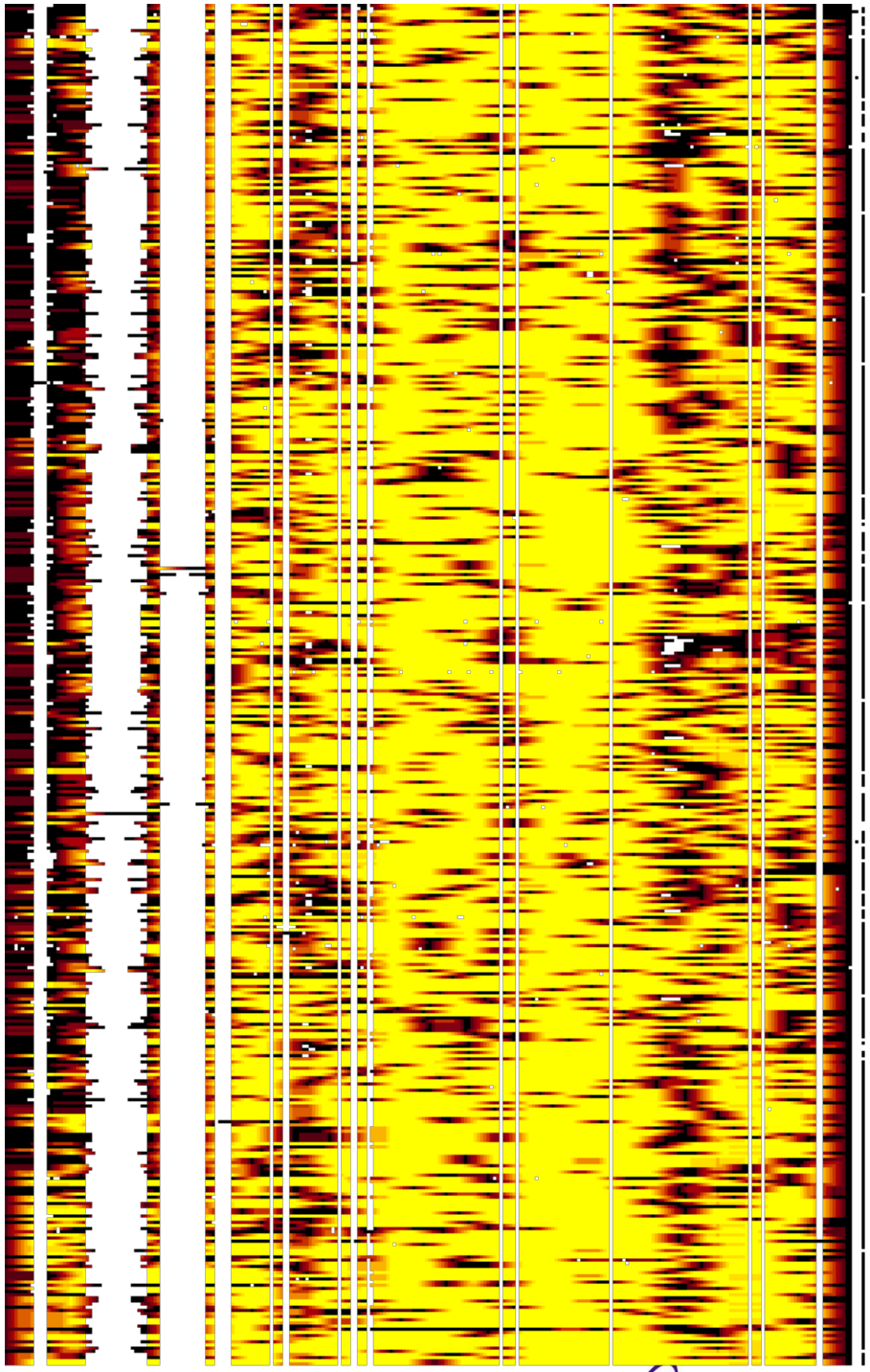
Each amino acid is colored according to the
set of 9-mers that contain it:

Yellow: all 9-mers that overlap with amino acid are perfectly matched in a test vaccine set;

Increasingly red: fewer and fewer matches in the overlapping set of 9-mers that span the amino acid;

Black: amino-acid residues that are not included in any vaccine set





Los Alamos
NATIONAL LABORATORY